

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:05:44 ; Search time 1.09218 Seconds
(without alignments)
1830.070 Million cell updates/sec

Title: US-09-806-703A-12

Perfect score: 74

Sequence: 1 QYIKANSKFIGITEL 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A Geneseq 101002.*

1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	74	100.0	15	11	AA06310
2	74	100.0	15	18	AAW35506
3	74	100.0	15	18	AAW11505
4	74	100.0	15	19	AAW67033
5	74	100.0	15	19	AAW71321
6	74	100.0	15	20	AAW04051
7	74	100.0	15	20	AAW67578
8	74	100.0	15	20	AAW73220
9	74	100.0	15	21	AAW45511
10	74	100.0	15	21	AAW82637
					Tetanus toxin epit
					Universal T-cell e
					Tetanus toxoid uni-
					Universal helper T
					T-Helper epitope f
					T-cell epitope pep
					Tetanus toxoid epi
					Tetanus P2 epitope
					Tetanus toxoid T c

11	74	100.0	15	21	AAW92625	Foreign epitope P2
12	74	100.0	15	21	AAW84427	Amino acid sequenc
13	74	100.0	15	21	AAW70300	Clostridium tetani
14	74	100.0	15	21	AAW44763	Tetanus toxoid pro
15	74	100.0	15	22	AAE11763	Clostridium tetani
16	74	100.0	15	22	AAW99515	Vaccine related MH
17	74	100.0	15	22	AAW85701	Amino acid sequenc
18	74	100.0	15	22	AAW85451	Wild-type T830 (t
19	74	100.0	15	22	AAW61956	Tetanus Toxoid uni
20	74	100.0	15	22	AAW20143	Tetanus toxin T-ce
21	74	100.0	15	22	AAW68636	HER-2 B cell pepti
22	74	100.0	15	22	AAW46172	Tetanus toxoid TT8
23	74	100.0	15	22	AAW49071	Tetanus toxoid TT
24	74	100.0	15	23	AAW97872	Tetanus toxin P2 (
25	74	100.0	16	18	AAW35445	T-cell stimulatory
26	74	100.0	16	20	AAW29705	Clostridium tetani
27	74	100.0	16	23	AAW93865	Clostridium tetani
28	74	100.0	16	23	AAW11413	Tetanus toxoid pre
29	74	100.0	17	15	AAW62692	Helper T cell epit
30	74	100.0	17	16	AAW82573	Tetanus toxin help
31	74	100.0	17	17	AAW05599	Tetanus toxin help
32	74	100.0	17	17	AAW88395	Tetanus toxin help
33	74	100.0	17	21	AAW93274	T-cell antigen TT2
34	74	100.0	17	21	AAW80856	HLA class II bindi
35	74	100.0	17	21	AAW54539	Pathogen derived T
36	74	100.0	17	21	AAW58768	T helper cell (Th)
37	74	100.0	17	22	AAW99516	Unidentified pepti
38	74	100.0	17	22	AAW62904	Vaccine related MH
39	74	100.0	17	22	AAW84435	Amino acid residu
40	74	100.0	17	22	AAW30941	Amino acid sequenc
41	74	100.0	17	22	AAW31029	Antigenic fragment
42	74	100.0	17	22	AAW31118	Antigenic fragment
43	74	100.0	17	22	AAW15589	Peptide 5 for pept
44	74	100.0	18	20	AAW26607	HIV-derived lipope
45	74	100.0	18	23	AAW809794	Peptide TT functio

ALIGNMENTS

RESULT 1
AAW06310
ID AAR06310 standard; protein; 15 AA.
XX
AC AAR06310;
XX
04-DSC-1990 (first entry)
XX
DE Tetanus toxin epitope.
XX
KW Tetanus toxin; vaccine; major histocompatibility complex; MHC;
KW antimalarial.
XX
OS Synthetic.
XX
PN EP378881-A.
XX
PD 25-JUL-1990.
XX
PF 27-DEC-1989; 89EP-0203318.
XX
PR 16-NOV-1989; 89IT-0022409.
PR 17-JAN-1989; 89IT-0019110.
XX
(ENIE) ENRICECHE SPA.
XX
PI Pessi A, Bianchi E, Verdini AS, Corradin G;
XX
DR WPI; 1990-225582/30.
XX
PT Synthetic peptide(s) corresp. to tetanus toxin epitope(s) - used
PT as universal carriers for prepn. of immunogenic conjugate(s) for
PT use as vaccines.

THIS PAGE BLANK (USPTO)

XX
PS
CC
CC
CC
CC
CC
CC
CC
CC
SQ

Claim 1; Page 17; 20pp; English.

Epitopic peptides may be used with synthetic haptens derived from a pathogen to generate an immune response to the pathogen. Peptides are recognised by numerous T-helper cell clones within the context of a wide range of alleles of the human MHC. The peptides may be used in an antimalarial vaccine inducing Ab. response to P.falciparum.

SQ Sequence 15 AA;

Query Match 100.0%; Score 74; DB 11; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.9e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QYIKANSKFIGITEL 15
DB 1 QYIKANSKFIGITEL 15

RESULT 2

AAW35506
ID AAW35506 standard; peptide; 15 AA.

XX AC AAW35506;

XX DT 22-APR-1998 (first entry)

XX DE Universal T-cell epitope peptide SEQ ID NO:8.

XX KW T-cell stimulatory peptide; immunogen; non-dendritic; carrier; tumour;
XX KW scaffold; inhibition; metastasis; wound healing; solid phase.

XX OS Unidentified.

XX FN WO9738011-A1.

XX PD 16-OCT-1997.

XX PF 03-APR-1997; 97WO-DE00146.

XX PR 03-APR-1996; 96DK-0000398.

XX PA (PEPR-) PEPRESEARCH AS.

XX PI Heegaard PMH, Jakobsen PH;

XX DR WPI; 1997-512645/47.

XX PT Non-dendritic peptide carrier linked to a solid phase - useful as a
XX PT diagnostic agent and as a scaffold for production of chemical
XX PT derivatives

XX PS Example 20; Page 124; 262pp; English.

XX CC A non-dendritic peptide carrier (A) has been developed which is coupled
XX CC through a linker to a solid phase, forming a complex of (A)-solid phase.
XX CC Where (A) comprises 10-50 amino acids capable of forming a secondary
XX CC structure in a benign buffer after liberation from the solid phase, and
XX CC further the (A)-solid phase complex comprises an immunogenic substance
XX CC and/or an immune mediator coupled on (A). The present sequence
XX CC represents a peptide used in an example from the present invention. An
XX CC (A)-solid phase complex can be used as a scaffold for the production of
XX CC chemical derivatives, characterised by covalently attaching molecules at
XX CC specific amino acid residues. Alternatively (A) is used as a scaffold-peptide for
XX CC incorporation into an immunostimulating Complex (Iscom) resulting in
XX CC an Iscom complex which is used for the chemical coupling of antigenic
XX CC peptides having fibronectin-, laminin- or vitronectin-like
XX CC domains. The complex is used for the promotion of cell-attachment to
XX CC particular to inhibit tumour growth and metastasis,
XX CC wound healing. Also a derivatised (A) can be used

CC
CC
CC
CC
CC
CC
SQ

for the selection of specifically-binding aptamers or as a diagnostic agent. Such diagnostic-(A) molecules could be used to detect molecules derived from or indicative of pregnancy or of a disease, such as an infectious, autoimmune or cancerous disease.

SQ Sequence 15 AA;

Query Match 100.0%; Score 74; DB 18; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.9e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QYIKANSKFIGITEL 15
DB 1 QYIKANSKFIGITEL 15

RESULT 3

AAW11505
ID AAW11505 standard; Protein; 15 AA.

XX AC AAW11505;

XX DT 24-SEP-1997 (first entry)

XX DE Tetanus toxoid universal Th epitope TT830.

XX KW Humanised antibody; anti-Fc receptor; H22; bifunctional; bispecific;
XX KW fusion protein; chimera; tetanus toxoid; helper T cell epitope;
XX KW antigen presentation; ds.

XX OS Clostridium tetani.

XX FN WO9640789-A1.

XX PD 19-DEC-1996.

XX PF 07-JUN-1996; 96WO-US09988.

XX PR 07-JUN-1995; 95US-0484172.

XX PA (MEDA-) MEDAREX INC.

XX PI Deo YM, Goldstein J, Graziano R, Somasundaram C;

XX DR WPI; 1997-052242/05.

XX DR N-PSDB; AAT58127.

XX PT Recombinant, multi-specific anti-Fc receptor antibody molecules -
XX PT also comprise an anti-target portion, used for the treatment of
XX PT cancer, autoimmune disease and pathogenic infection

XX PS Example 7; Fig 24; 115pp; English.

XX CC Synthetic DNA coding for the wild-type universal Th epitope from
XX CC tetanus toxoid, designated TT830, was fused to the 3'-end of DNA
XX CC encoding heavy chain sequences from the humanised anti-Fc gamma RI
XX CC monoclonal antibody H22. The resulting fusion protein was shown to
XX CC be significantly more efficient in antigen presentation and T cell
XX CC stimulation than the TT830 epitope alone. A similar fusion
XX CC construct was prepared coding for a mutant, antagonistic form of the
XX CC epitope (designated TT833S) fused to the anti-Fc gamma RI. The
XX CC Fab22-TT833S is at least 100 times more effective than TT833S in
XX CC inhibiting T cell activation.

SQ Sequence 15 AA;

Query Match 100.0%; Score 74; DB 18; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.9e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QYIKANSKFIGITEL 15

DB 1 QYIKANSKFIGITEL 15

THIS PAGE BLANK (USPTO)

03:12 2003

DT 26-NOV-1998 (first entry)
XX Universal helper T-cell epitope P2 derived from tetanus toxin.
DE
XX
XX Liver stage; Plasmodium; Navy Yoelii Liver Stage 3 antibody; NYLS3;
KW hepatic and erythrocytic stage protein; PyHEP17; vaccine;
KW malaria parasite; tetanus toxin; P2; helper T-cell epitope.
XX
XX Synthetic.
OS Clostridium tetani.
XX
XX US5814617-A.
PN
XX
XX 29-SEP-1998.
PD
XX
XX 07-OCT-1994; 94US-0319704.
PF
XX
XX 07-OCT-1994; 94US-0319704.
PR
XX
XX (USNA) US SEC OF NAVY.
PA
XX Charoenvit Y, Doolan DL, Hedstrom RC, Hoffman SL;
PI
XX WPI; 1998-541794/46.
DR
XX
XX Vaccine for protecting mammal against infection by malaria caused by
PT Plasmodium species - comprises a first nucleic acid encoding a first
PT polypeptide capable of eliciting an immune reaction against an
PT antigen expressed during the liver
XX
XX Disclosure; Column 12; 24pp; English.
XX
XX AA71321-22 represent universal helper T-cell epitopes derived from
CC tetanus toxin. They are used to enhance host immune response to
CC vaccines. The specification describes a Plasmodium yoelii liver stage
CC 17 kDa hepatic and erythrocytic stage protein designated PyHEP17. This
CC protein elicits a response from an Igi monoclonal antibody designated
CC Navy Yoelii Liver Stage 3 (NYLS3). This antibody does not recognise
CC sporozoites, but does recognise P. yoelii liver stage parasites. NYLS3
CC eliminates up to 90% of liver stage parasites. The specification
CC describes a vaccine for reducing the severity or incidence of infection
CC by a malaria parasite of the genus Plasmodium. The DNA vaccine comprises
CC exon 1 and part of exon 2 of the PyHEP17 gene.
XX
XX SQ Sequence 15 AA;
Query Match 100.0%; Score 74; DB 19; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.9e-07; Indels 0; Gaps 0;
Matches 15; Conservative 0; Mismatches 0;
QY 1 QYIKANSKFIGITEL 15
DB 1 QYIKANSKFIGITEL 15
RESULT 6
AAW04051
ID AAW04051 standard; peptide; 15 AA.
XX
XX AC AAW04051;
XX
XX 04-JAN-2000 (first entry)
DT
XX
XX T-Helper epitope from tetanus toxoid.
DE
XX
XX Covalently reactive antigen analog; CRAA; catalytic antibody;
KW electrophilic reaction centre; phosphonate; boronate; vaccine;
KW transition state analog; TSA; isostere; gp120; HIV-1; T-helper;
KW tetanus; toxoid; B-T-epitope.
XX
XX OS Clostridium tetani.
XX
XX PN WO9948925-A1.

DT 15-DEC-1998 (first entry)
XX
XX AAW67033 standard; peptide; 15 AA.
AC
XX
XX AAW67033;
DT
XX
XX Tetanus toxin fragment (residues 830-844).
DE
XX
XX Tetanus toxin; vaccine; antibody; carbohydrate peptide conjugate;
KW dendrimeric poly-lysine; epitope; tumour.
KW
XX
XX Clostridium tetani.
OS
XX
XX WO9843677-A1.
PN
XX
XX 08-OCT-1998.
PD
XX
XX 27-MAR-1998; 98WO-EP01922.
PF
XX
XX 27-MAR-1997; 97US-0041726.
PR
XX
XX (INSP) INST PASTEUR.
PA
XX Bay S, Cantacuzene D, Leclerc C, Lo-man R;
PI
XX WPI; 1998-557071/47.
DR
XX
XX Carbohydrate peptide conjugate used as vaccine - comprises carrier
PT with dendrimeric poly-lysine enabling multiple epitopes to be
PT covalently attached
PT
XX
XX Disclosure; Page 13; 55pp; English.
XX
XX The invention relates to a new carbohydrate peptide conjugate, which
CC comprises a carrier with a dendrimeric poly-lysine enabling multiple
CC epitopes to be covalently attached to it. Also claimed are: (1) an
CC antibody purified from biological fluid or cells of organisms
CC administered with the carbohydrate peptide conjugate, and (2) a diagnosis
CC kit comprising antigen-specific antibodies elicited by immunisation with
CC the carbohydrate peptide conjugate. The peptide conjugate, antibody and
CC diagnosis kit are used to provide pharmaceutical compositions and
CC vaccines against tumours. These can be used to support an immune response
CC against viral infections caused by hepatitis virus, HIV or cytomegalovirus.
CC They can be used to enhance immune responses, especially B- and T-cell responses, of humans and animals against bacterial infections. The
CC carbohydrate peptide conjugate stimulates the antibody and T-cell
CC response without stimulating undesired immune responses. The composition
CC is capable of increasing the survival of tumour bearing humans and
CC animals. The present sequence corresponds to residues 830-844 of tetanus
CC toxin. The synthetic peptide corresponding to this sequence may be used
CC as an epitope in a carbohydrate peptide conjugate.
XX
XX SQ Sequence 15 AA;
Query Match 100.0%; Score 74; DB 19; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.9e-07; Indels 0; Gaps 0;
Matches 15; Conservative 0; Mismatches 0;
QY 1 QYIKANSKFIGITEL 15
DB 1 QYIKANSKFIGITEL 15
RESULT 5
AAW71321
ID AAW71321 standard; peptide; 15 AA.
XX
XX AC AAW71321;
XX

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:05:44 ; Search time 1.52905 Seconds
(without alignments)
1830.070 Million cell updates/sec

Title: US-09-806-703A-14
Perfect score: 112
Sequence: 1 FNNFTVSWLRVVKVSASHLE 21

Scoring table: BLOSUM62
Gapop 10.0 ; Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	112	100.0	21	AA11896	Immunogenic conjug
2	112	100.0	21	AA11896	Tetanus toxoid pro
3	112	100.0	21	AA11896	T-cell antigen TT3
4	112	100.0	21	AA11896	Broad range helper
5	112	100.0	21	AA11896	Tetanus toxin frag
6	112	100.0	21	AA11896	T-cell epitope pep
7	112	100.0	21	AA11896	Tetanus toxoid epi
8	112	100.0	21	AA11896	Tetanus p30 epitop
9	112	100.0	21	AA11896	Tetanus toxin T ce
10	112	100.0	21	AA11896	Foreign epitope P3

11	112	100.0	21	AA11896	Amino acid sequenc
12	112	100.0	21	AA11896	CD4+ T cell epitop
13	112	100.0	21	AA11896	Clostridium tetani
14	112	100.0	21	AA11896	Amino acid sequenc
15	112	100.0	21	AA11896	Universal tetanus
16	112	100.0	21	AA11896	Tetanus Toxoid uni
17	112	100.0	21	AA11896	Tetanus Toxoid T-ce
18	112	100.0	21	AA11896	HER-2 B cell pepti
19	112	100.0	21	AA11896	Tetanus toxoid TT9
20	112	100.0	21	AA11896	Tetanus toxoid TT
21	112	100.0	21	AA11896	Tetanus toxoid pre
22	112	100.0	21	AA11896	Tetanus toxoid 947
23	112	100.0	21	AA11896	PSMpep010 - P30 in
24	112	100.0	21	AA11896	PSMpep011 - P30 in
25	112	100.0	21	AA11896	PSMpep012 - P30 in
26	112	100.0	21	AA11896	LHRH-containing im
27	112	100.0	21	AA11896	Amyloid beta/tetan
28	112	100.0	21	AA11896	Synthetic immunoge
29	112	100.0	21	AA11896	Peptide comprising
30	112	100.0	21	AA11896	A peptide which ma
31	112	100.0	21	AA11896	Synthetic immunoge
32	112	100.0	21	AA11896	Tetanus toxoid 830
33	112	100.0	21	AA11896	Amyloid beta/tetan
34	112	100.0	21	AA11896	Tetanus toxoid epi
35	112	100.0	21	AA11896	Amyloid beta/tetan
36	112	100.0	21	AA11896	Synthetic immunoge
37	112	100.0	21	AA11896	Tetanus toxoid epi
38	112	100.0	21	AA11896	Amyloid beta/tetan
39	112	100.0	21	AA11896	Peptide comprising
40	112	100.0	21	AA11896	A peptide which ma
41	112	100.0	21	AA11896	Immunogenic branch
42	112	100.0	21	AA11896	Immunogenic branch
43	112	100.0	21	AA11896	Immunogenic branch
44	112	100.0	21	AA11896	Tetanus toxoid epi
45	112	100.0	21	AA11896	Tetanus toxoid epi

ALIGNMENTS

RESULT 1
ID AA11896 standard; peptide; 21 AA.
XX AA11896;
AC AA11896;
XX 19-JUL-1991 (first entry)
DT Immunogenic conjugate constituent peptide, TT3.
DE Immunogenic conjugate constituent peptide, TT3.
XX Malaria vaccine; major histocompatibility complex.
KW Tetanus toxin.
OS Tetanus toxin.
PH Key Location/Qualifiers
FT Peptide 1..14
FT /label= active fragment (claimed)
XX EP427347-A.
XX 15-MAY-1991.
PD 07-NOV-1990; 90EP-0202948.
XX 10-NOV-1989; 89IT-0022355.
XX (ENIE) ENIRICERCH SPA.
XX Bianchi E, Pessi A, Corradin G;
XX WPI; 1991-141874/20.
XX Synthetic peptide(s) used as universal carriers - for preparing

THIS PAGE BLANK (USPTO)

Tue Jul 22 12:43:19 2003

us-09-806-703a-14.rag

PT immunogenic conjugates used as vaccines against Plasmodium
PT falciparum
XX
PS Claim 1; page 13; 16pp; English.
XX
CC This peptide corresps. to residues 947-967 of Tetanus toxin. It can
CC be used as a universal carrier for the prepn. of an immunogenic
CC conjugate. It is covalently bound to a peptide or polysaccharide
CC hapten derived from a pathogen. This conjugate can be used as
CC a vaccine for malaria. This peptide is recognised by different T-
CC helper cell clones in association with alleles of the human MHC.
CC It contains 2 epitopes: (a) 953-967, recognised by DR5-restricted
CC clones; and (b) 947-960, recognised by all other DR and DP-
CC restricted clones.
XX
SQ Sequence 21 AA;

Query Match 100.0%; Score 112; DB 12; Length 21;
Best Local Similarity 100.0%; Pred. No. 3.9e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FNNFTVSFWLRVPKVSASHLE 21
Db 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 2
AAW06130
ID AAW06130 standard; Peptide; 21 AA.
XX
AC AAW06130;
XX
DT 07-FEB-1997 (first entry)
XX
DE Tetanus toxoid protein T-cell epitope.
XX
KW Cholesteryl ester transfer protein; CETP; antigen; vaccine;
KW cardiovascular disease; atherosclerosis; tetanus toxoid;
KW T-cell epitope.
XX
OS Clostridium tetani.
XX
PN WO9634888-A1.
XX
PD 07-NOV-1996.
XX
PF 01-MAY-1996; 96WO-US06147.
XX
PR 01-MAY-1995; 95US-0432483.
XX
PA (TCEL-) T CELL SCI INC.
XX
PI Rittershaus CW, Thomas LJ;
XX
DR WPI; 1996-506103/50.
XX
PT Cholesteryl ester transfer protein B cell epitope linked to T cell
PT epitope - used to generate vaccine to regulate CETP activity for
PT decreasing the risk of developing a cardiovascular disease e.g.
PT atherosclerosis
XX
PS Claim 11; Page 43; 72pp; English.

XX
CC A helper T-cell epitope (AAW06130) comprises amino acids 947-967
CC of tetanus toxoid protein. It can be utilised in novel peptide
CC vaccines (see also AAW06129, AAW06132) also including B-cell
CC epitopes from human or rabbit cholesteryl ester transfer
CC protein. This peptide can induce an immune response against endogenous
CC proteins or preventing a cardiovascular disease.

Query Match 100.0%; Score 112;
Best Local Similarity 100.0%; Pred. No.
Matches 21; Conservative 0; Mismatch

Qy 1 FNNFTVSFWLRVPKVSASHLE 21
Db 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 3
AAR88397
ID AAR88397 standard; Peptide; 21 AA.
XX
AC AAR88397;
XX
DT 12-JUN-1996 (first entry)
XX
DE T-cell antigen TT3 peptide.
XX
KW T-antigen; vaccine; antibody; T-cell; T-
KW alpha-helix; coiled-coil heterodimer; cc
XX
OS Synthetic.
XX
PN WO9531480-A1.
XX
PD 23-NOV-1995.
XX
PF 18-MAY-1995; 95WO-CA00293.
XX
PR 18-MAY-1994; 94US-0245507.
XX
PA (SPIS-) SPI SYNTHETIC PEPTIDES INC.
XX
PI Cachia PJ, Hodges RS, Houston ME, Irv
PI Zhou NE;
XX
DR WPI; 1996-010880/01.
XX
PT Hetero:dimeric polypeptide immunogen in
PT with different antigens on each sub:unit
PT for antibody prodn.
XX
PS Claim 7; Page 62; 95pp; English.
XX
CC This T-cell antigen TT3 peptide may be
CC contained in one of the 2 subunits of ar
CC heterodimer. Each core peptide is comp
CC internal AA repeat sequences. This pept
CC to the core peptide through covalent lin
CC internal repeats. The 2 subunits of the
CC in a stable alpha-helical coiled-coil co
CC stoichiometry, and the peptide antigen i
CC surfaces of the configuration. The hete
CC synthetic vaccine (optionally multivaler
CC antibodies.
XX
SQ Sequence 21 AA;

Query Match 100.0%; Score 112;
Best Local Similarity 100.0%; Pred. No.
Matches 21; Conservative 0; Mismatch

Qy 1 FNNFTVSFWLRVPKVSASHLE 21
Db 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 4
AAW46449
ID AAW46449 standard; Peptide; 21 AA.
XX
AC AAW46449;

THIS PAGE BLANK (USPTO)


```

Db 1190 SASSG-----PPKAEEDYVNEPLYNTFTNALGNAEYMKNSLLSVPEKAKAFDNDYW 1243
Qy 1230 QDDPPERGA--PPSTFKGTPT-----AENPEYL 1255
Db 1244 NLSLPPRSTLQHPDYLOEYSTKYFYKQNGRIRPIVAENPEYL 1285

RESULT 8
XMRK XIPMA
ID XMRK XIPMA STANDARD; PRT; 1167 AA.
AC P13388;
DT 01-JAN-1990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
GN XMRK OR TU.
OS Xiphophorus maculatus (Southern platyfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX NCBI_TaxID=8083;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90015140; PubMed=2797166;
RA Wittbrodt J., Adam D., Malitschek B., Maueier W., Raulf F.,
RA Telling A., Robertson S.M., Scharlt M.;
RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
RL inducing Tu locus in Xiphophorus.";
RN Nature 341:415-421(1989).
RN [2]
RP REVISION TO 515.
RA Scharlt M.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC 1- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
CC 1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC 1- SUBCELLULAR LOCATION: Type I membrane protein.
CC 1- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
CC 1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; X16891; CAA34770.2; -.
CC PIR; S06142; S06142.
CC HSSP; P11362; 1FGK.
CC InterPro; IPR000494; EGFR_L domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR002290; Ser_thr_pkinase.
CC InterPro; IPR01245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF01030; Recep_L domain; 2.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 5.
CC SMART; SM00220; S_TKc; 1.
CC SMART; SM00219; Tyrc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.
CC PROSITE; PS00109; PROTEIN KINASE TYR; 1.
CC PROSITE; PS50011; PROTEIN KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
FT SIGNAL 1
FT CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE

```

```

FT DOMAIN 26 642
FT TRANSMEM 643 665
FT DOMAIN 666 1167
FT DOMAIN 710 977
FT NP_BIND 716 724
FT BINDING 743 743
FT ACT_SITE 835 835
FT DISULFID 195 204
FT DISULFID 220 228
FT DISULFID 224 236
FT DISULFID 237 245
FT DISULFID 251 253
FT DISULFID 256 265
FT DISULFID 269 296
FT DISULFID 300 311
FT DISULFID 315 330
FT DISULFID 333 337
FT DISULFID 504 513
FT DISULFID 508 521
FT DISULFID 524 533
FT DISULFID 537 553
FT DISULFID 556 569
FT DISULFID 560 577
FT DISULFID 593 615
FT DISULFID 618 626
FT DISULFID 622 634
FT CARBOHYD 114 114
FT CARBOHYD 144 144
FT CARBOHYD 201 201
FT CARBOHYD 356 356
FT CARBOHYD 365 365
FT CARBOHYD 398 398
FT CARBOHYD 417 417
FT CARBOHYD 501 501
FT CARBOHYD 576 576
FT CARBOHYD 621 621
SQ SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;

Query Match 38.78; Score 2649.5; DB 1; Length 1167;
Best Local Similarity 44.68; Pred. No. 6.7e-134;
Matches 570; Conservative 164; Mismatches 390; Indels 155; Gaps 30;

Qy 4 AALRCWGLLALLPPGAAS-----QVCTGTDMLKRLPASPEHLDMLRHLYQGCVVQGN 59
Db 8 AALLQ--LLLVLSIRCCSTDPRKVCQGTSNQMTM---LDNHYLKKMKWYSGGVNLEN 62
Qy 60 LEITYLPTNASLSFLQDIQEV-----QFNFTVSFWLVPKVSASHLEQRLRVRGTQL 113
Db 63 LEITYTQENQDLSFLQSIQEVGGVLIAMNEVST-----IPLVN-----LRLIRGNL 110
Qy 114 FEDNYALAVLDNGDPLNNTTPTVGASPGGLRELQLSLTEILKGVLIQRNPOLCYODTI 173
Db 111 YEGNFTLLVMSNYQK-PPSSP--DVYQVGLKQLQSLNLTLEILSGGVKVSHPNLLCNVETI 167
Qy 174 LWKDIFHKNQALTLTIDNRSRACHPCSPMKSGRCWGSESDCQSLTRTVCAAGC-AR 232
Db 168 NWWDIVDKTSNPTNMLIPHAFERCQCKDHCVCNGSCWAPGPGHCQFTELKLCRAQCNR 227
Qy 233 CKGFLPTDCCHQCAAGCTGPKHSDCILACLFHNSGICELHCPALVTYNTDTFSPMPNE 292
Db 228 CRGPKPIDCCNEHCAGCTGPRATDCLACRDFNDGDKTCTCPPPKIYDIVSHQVVDNPN 287
Qy 293 GRITFGASCVTACPNYVLSTDVSGCTLVCPHMQEVTAEADGTORCEKCSKPCARVCVGLG 352
Db 288 IKYTFGAACVKECPNSVVVTE-GACVRSAGMLEVD-ENGKRSCKPCDGVCPKVCDDIG 345
Qy 353 MEHLREVRAVTSANIOEFACKKIFGSLAFLPESFEDGPRASNTAPLOPEOLQVPELSEI 412
Db 346 IGSLSNTIAVNSTNIRSFNCTKINGDIILNRNSFEGDPHYKIGTMDPEHLWNLTUTVKEI 405
Qy 413 TGYLYISAWPDSLPDLSVFQNLQVIRGILHNGAYS--LTQLGLGISWLRLSRELGSGL 471

```

```
Db 406 TGVLVIMWMPENMTSLVFQNLRIIRGTTTSRGFSFVVVQVRLQWLGLASLKEVSAGN 465
Qy 472 ALIHNTLFCVHTVPNDQLRPNHQALHTANRPEDCEVGEGLACHOLCARGHCWPGP 531
Db 466 VILKNTLQLRVANTIRNRRLSRSDQSTIEYDART-----ENQTCNNECEDGCGWPGP 518
Qy 532 TQCYNCSQFLRGQCEVCRVLQGLPREYVVARHCLPCHPCQCPQNGSVTCFGEADQCV 591
Db 519 TMCVSLHVDGRGRCVASCNLQGEPREAQVDGRVCVQCHQCLVQTSLTCTCYGPGPANC 578
Qy 592 ACAHYKDPFPFCVAPCPGKEDLSYMPITWKFPEDEGACQPCFINCTHSCVDLDDKGPFAE 651
Db 579 KSAHFQDGPQICPRCPHGILGDGTL-LWKYADRWGQCQPCQNCQTCGSGPLSGCRGD 637
Qy 652 QRASPLTSIVSAVGVILLVVLGVVFGILIKRQOKIRKYMRRLLQBTVELVBLTSGA 711
Db 638 -IVSHSLAVGLVSGLLITVIVALLIVVLLRRRIK-RKRTIRCLQBEKELVEPLTSGQ 695
Qy 712 MPNOAMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANK 771
Db 696 APNQAFLIKETEFKORVLGSGAFGVYKGLWNPDPGENIRIPVAIKVLRENTSPKANK 755
Qy 772 EILDEAYVMAGVSPYSRLGICLTSTVQLVTQMLPYGCLLDHVRENRGLSGQDLNW 831
Db 756 EVLDEAYVMSVDHPVHCRLGICLTSAVQLVTQMLPYGCLLDVYRQHQERICQGLNW 815
Qy 832 CMQIAKMSYLEDVRLVHRDLAARNLVKSNHVKITDFGLARLLDIDETHYHADGGKVP 891
Db 816 CVQIAKGMNLEERHLVHRDLAARNLVKSNHVKITDFGLSKLLTADKEYQADGGKVP 875
Qy 892 IKWMALESILRRRTHQSDVMSYGVYVWELMTFGAKPYDGPAREIPDLLEKGERLPPOP 951
Db 876 IKWMALESILQWYTHQSDVMSYGVYVWELMTFGSKPYDGPAREIPDLLEKGERLPPOP 935
Qy 952 ICTIDVYIMVKWMIDSECPRELFSEFSEMRARDPQRFVVIQNEQDLGPASPLDSTFY 1011
Db 936 ICTIEVYIMILKWMIDSPSRPRELVGEFSQWARDPSRYLVQ--NLPSLSDRFLP 992
Qy 1012 RSLLEDDMDGLVDAEYLVQOQFFCPDPAPAGGWWHRRSSSTRSGGDLTLGLPEP 1071
Db 993 SRLLSDD--DVNDAEYLLPKYI-----NRQGS----- 1020
Qy 1072 SEEAAPSLAPSEAGSDVFDGDLGMAAGLQSLPHTDPSLQRYSEDPTV-PLPSET 1130
Db 1021 -----EPCIPPTGH-----PVRENSITLRNDSDFQNALEKDL 1053
Qy 1131 DGYVAPLTCSPQEVYVQDVRPOP-----PSPRE-----GPLP-AARPAGATLERAK 1177
Db 1054 DGH-----EYVNPQGETSSRLSDIYNPNVEDLTDGWPVSLSSQEAETNFSRPE 1103
Qy 1178 TSLPGKNGVXKDVAFGAGAVENPEYLTPOGGAAPHPPPAFSADFNLVWDQDPBPG 1237
Db 1104 YLNTQNLSL---PLVSSGSMDDPDY---QAG-----YQAAF-----LPQTG 1138
Qy 1238 APPSTFKTPTAENPEYLG 1256
Db 1139 ALTGNMGFLPAENLEYLG 1157
```

RESULT 9

ERB3 HUMAN

ID ERB3 HUMAN STANDARD; PRT: 1342 AA.

AC P21860;

DT 01-MAY-1991 (Rel. 18, Created)

DT 01-MAY-1991 (Rel. 18, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)

DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).

CN ERB3 OR HER3

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI_TaxID=9606;

[1]
SEQUENCE FROM N.A.
MEDLINE=90093234; PubMed=2687875;
Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
"Isolation and characterization of ERBB3, a third member of the
ERBB/epidermal growth factor receptor family: evidence for
overexpression in a subset of human mammary tumors.";
Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).

[2]
SEQUENCE FROM N.A.
MEDLINE=90311312; PubMed=2164210;
Plozman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
Todaro G.J., Shoyab M.;
"Molecular cloning and expression of an additional epidermal growth
factor receptor-related gene.";
Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909 (1990).

[3]
SEQUENCE FROM N.A. (SHORT FORM).
TISSUE=Placenta;
MEDLINE=93282822; PubMed=7685162;
Katoch M., Yazaki Y., Sugimura T., Terada M.;
"c-erbB3 gene encodes secreted as well as transmembrane receptor
tyrosine kinase.";
Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
-1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTKA.
-1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
-1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
(POTENTIAL).
-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
SECRETED (SHORT FORM).
-1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING..
-1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
-1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
-1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
-1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
-1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M29366; AAA35790.1; --
EMBL; M34309; AAA35979.1; --
EMBL; S61953; AAB26935.1; --
PIR; A36223; A36223.
HSP; P11362; 1FGK.
Genew; HGNC:3431; ERBB3.
MIM; 190151; --
InterPro; IPR000494; EGFR_L_domain.
InterPro; IPR00719; Euk_pkinase.
InterPro; IPR002174; Furin-like.
InterPro; IPR001245; Tyr_pkinase.
Pfam; PF00069; pkinase; 1.
Pfam; PF00757; Furin-like; 1.
Pfam; PF01030; Recep_L_domain; 2.
ProDom; PD000001; Euk_pkinase; 1.
SMART; SM00261; FU; 3.
SMART; SMO0219; TyKc; 1.
PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
PROSITE; PS00119; PROTEIN_KINASE_TYR; FALSE_NEG.
PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;

[illegible]

```
RESULT 10
ERB3_RAT
ID ERB3_RAT STANDARD; PRT; 1339 AA.
AC Q62799; O62955;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (C-erbB3).
GN ERB3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=96096535; PubMed=8522190;
RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
RT "Cloning of the rat ErbB3 cDNA and characterization of the
RT recombinant protein.";
RL Gene 165:279-284 (1995).
RN [2]
RP REVISIONS TO 85; 513 AND 565.
RA Hellyer N.J., Koland J.G.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 922-1097 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659 (1997).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTK.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U29339; AAC28498.2; -
DR EMBL; U52530; AAC3050.1; -
DR HSSP; P11362; IFGK.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; FALSE NEG.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; signal;
```

```
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 19
FT CHAIN 20 1339
FT DOMAIN 20 643
FT TRANSMEM 644 662
FT DOMAIN 663 1339
FT DOMAIN 183 259
FT DOMAIN 707 964
FT NP_BIND 713 721
FT BINDING 740 740
FT ACT_SITE 832 832
FT DISULFID 196 194
FT DISULFID 190 202
FT DISULFID 210 218
FT DISULFID 214 226
FT DISULFID 227 235
FT DISULFID 231 243
FT DISULFID 246 255
FT DISULFID 259 286
FT DISULFID 290 301
FT DISULFID 305 320
FT DISULFID 323 327
FT DISULFID 500 509
FT DISULFID 504 517
FT DISULFID 520 529
FT DISULFID 533 549
FT DISULFID 556 573
FT DISULFID 576 585
FT DISULFID 613 621
FT DISULFID 617 629
FT CARBOHYD 126 126
FT CARBOHYD 250 250
FT CARBOHYD 353 353
FT CARBOHYD 408 408
FT CARBOHYD 414 414
FT CARBOHYD 437 437
FT CARBOHYD 469 469
FT CARBOHYD 522 522
FT CARBOHYD 566 566
FT CARBOHYD 616 616
FT CONFLICT 1028 1028
SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F2402BBDF1E CRC64;
Query Match 33.8%; Score 2316.5; DB 1; Length 1339;
Best Local Similarity 40.4%; Pred. No. 4.3e-116;
Matches 523; Conservative 170; Mismatches 429; Indels 173; Gaps 37;
Oy 3 LAALCRWGLLLALLPPGAA---STQVCTGTDMKRLPASPETHLDMLRHLVYQGVVQGN 59
Db 7 LQVLC----FLLSLARGSEMGNSQAVCPGTLNGLSVTGDADNOYQTLVLYKEVWGN 62
Oy 60 LEITYLPTNASLSFLQDIQE-----VQFNFTVSWFLRVKPVKVSASHLEORLIRVGTQL 113
Db 63 LEIVLTGHNADSLFLQIREVTGVVLMVAMNEFSV---LPLP-----NLRVVRGTQV 110
Oy 114 FEDNYALAVLDNGDPINNTPVTGASPGGLRELQLRSITELKGGVLQIRNPOLCYODTI 173
Db 111 YDGKFAIFVM-----LNYNT-----NSSHALRQLKFTQTLTEILSGGVYIERNDKLCHMDTI 161
Oy 174 LWKDIFHKNNQLALTLDITNRSRACHPCSPMKSCRCWGESSEDCQSILRTVTCAGGC-AR 232
Db 162 DWRDIVRVR---GAEIVVKNNGANCPCHEVCCKG-RCWGPGPDDCOLITKTICAPQCNGR 217
Oy 233 CKGELPTDCCHEQCAAGCTGPKSHDCLACLHFNHSGICELHCPALVYTYNTDTFFSMNPDE 292
Db 218 CFGENPNQCCHDECAGCGSGPQDTCFACRFNDSGACVPRCPPLVYNNKLTFLQLEPNPH 277
Oy 293 GRITFGASCVTACPNYLVSTDVGSCITLVCLPHNQEVTAEDGTQRCCKSKCARVCVCLG 352
Db 278 TKYQYGGVGVASCAPHNFV-VDQTFVCRACPPDKMEVD-KHGLKMKCEPCGGGLCPKACEGTG 335
Oy 353 MEHLREVRVTSANIQEFAGCKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFTLEBEI 412
```

Db 336 SG--SRVQTVSSNIDGVNCTKILGNLDFLITGLNVDPMHKKIPALDPEKLNFRFTVREI 393
Qy 413 TGYLYISAWPDLPLDLSFQNLQVIRGRIILHNGAYS-LTLQGLGISMGLRSLRELGSGL 471
Db 394 TGYLNIQSWPHMNFVSFNLITIGESLYNRGFSLLIMKNLNVTSLGFRSLKEISAGR 453
Qy 472 ALIHNTHLCPVTVFMDQLPFRNHQALLHTA-NRPDECEVGEGLACHQLCARGHCWGP 530
Db 454 VYISANQOLCYHHSNLWTRLLRGPSEERLDIKYDRPLGECLEAGKVCDPGLCSSGCGWGP 513
Qy 531 PTQCVCNSQFLRGCEVEECVLOGLPREYVYNARHCLPCHPECOPNGSVTCFGEADQC 590
Db 514 PGQLSCRNSYREGVCTVCHNCFLOGEPRFVHEAQCFSCHECLPFMEGTSTCNGSGSDAC 573
Qy 591 VACAHYKDPFPCVARGSGVKPDLSPYIMKFFPDEEGACQPCINCTHSC--VDLDDKGC 648
Db 574 ARCAFRDGPCHVNSCPHGILG--AKGPIYKYPDAQNECRCHENCTGCGNGPQLDCLG 631
Qy 649 PAQRASPLTSIVSAVGIILVVLGVVFGVILIKRQOKIR-KYTMRLRLLOETELVEPLT 707
Db 632 QAEVLMKPHLVIAVTVG--LAVILMILGGSFLYWRGRIQNKAMRRYLERGESIEPLD 689
Qy 708 PSGAMPQAQMRILKETELRKVKVLGSCAFGVYKGIWIPGENVKIPVAIKVLENTSP 767
Db 690 PS-EKANKVLARIKTELRLKVLGSGVFGVTHKGIWIPGESIKIPVCIKVIEDKSGR 748
Qy 768 KANKEIIDEAVYVAGVSPYVRLILCTVLTOLMPYVGLDHLVHNRNRLGSGD 827
Db 749 QSFQAVTDHMLAVGSLDHAHIVRLILGLCPGSLQLVLYPLGSLDLHVAKHRETLGPOL 808
Qy 828 LLNWCMIQAKMSYLEDLVRLVHRLDAARNLVKSPNHVKIDFLGLARLLDIDETEHADG 887
Db 809 LLNMGVQIAKMYLEHSMVHRDLARNVLMKSPSQVQVADFLPDDQLLHSE 868
Qy 888 GKVPDKMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERL 947
Db 869 AKTPIKMALESIHFGKYTHQSDVMSYGVTVWELMTFGAEFYAGLRLAEIPDLLEKGERL 928
Qy 948 POPPICTIDVYIMVKCMIDSECRPRELVSEFSRMARDPQRFVWIONEDLGPASPLD 1007
Db 929 AQPOICTIDVYIMVKCMIDENIRPTFKELANEFTRWARDPPRYLVIKRAS-GFGTP-- 985
Qy 1008 STFYRSLLDDMDGLVDAEYLVPOQGFCCPDPAAGAGGMVHHRSSSTRSGGDLTL 1067
Db 986 PAAPSVLTTEL-----QAELEPEL-----DLDL 1011
Qy 1068 GLEPSEE-----EAPRSLAPSEG-----AGSDVFGDLOMG 1099
Db 1012 DLEAEEGLATSLGSLSLPTGTLTRPRGSQSLSPSSGYMPMNQSSLGEACLDSAVLGG 1071
Qy 1100 AAKGLQSLPHTDPLQRYSEDPTVPLPSETDGVY---APL-----TC-----SPQP 1143
Db 1072 REQSRPISLH-PIPRG-----PASESEGHVTSGEAELOEKSVVCRSRSRSPRP 1123
Qy 1144 E-----YVNPQDVRPQPSPREGP-----LPAARPAGATLERAKTLSP-GKNGV 1187
Db 1124 RGDAXHSQRHSLTPTVPLSPGLLEEDNGYVMPDTHLKGASSREGTLLSSVGLSSVL 1183
Qy 1188 -----KDVFAFGAVENPEYLTPOGGNAOPHP 1216
Db 1184 GTEEEDE-----BEYEYMNKRGRSP-PRPP 1209

RESULT 11
EGFR_DROME STANDARD; PRT; 1426 AA.
AC P0412; O61601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Bgrf)
DE (Gurken receptor) (Torpedo protein) (Drosophila relative of ERBB).

GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM TYPES I AND II).
RX MEDLINE=94350209; PubMed=8070664;
RA Clifford R., Schupbach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT that several genetically defined classes of alleles cluster in
RL subdomains of the receptor protein.";
RL Genetics 137:531-550(1994).
RN [2]
RP REVISIONS.
RA Clifford R., Schupbach T.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Livneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RT hormone binding and kinase domains.";
RL Cell 40:599-607(1985).
RN [4]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RX STRAIN=Oregon-R; TISSUE=Embryo;
RL MEDLINE=870202474; PubMed=3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RT Drosophila EGF receptor homolog transcripts.";
RL Cell 46:1091-1101(1986).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RP ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
RT specification in wild-type, Ellipse, and null mutant Drosophila.";
RL Dev. Biol. 205:129-144(1999).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RX STRAIN=Berkley;
RL MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler K.C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of *Drosophila melanogaster*";
RL Science 287:2185-2195(2000).
RN [7]
RN SEQUENCE OF 959-1078 FROM N.A.
RC STRAIN=Daekwanryeong;
RX MEDLINE=85137938; PubMed=2983232;
RA Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
RT "A *Drosophila* genomic sequence with homology to human epidermal
RT growth factor receptor";
RL Nature 314:178-180(1985).
RN [8]
RN SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
RP ANALYSIS.
RX MEDLINE=92038942; PubMed=1936959;
RA Raz E., Schejter E.D., Shilo B.Z.;
RT "Interallelic complementation among DER/flb alleles: implications for
RT the mechanism of signal transduction by receptor-tyrosine kinases";
RL Genetics 129:191-201(1991).
RN [9]
RN REVIEW.
RP MEDLINE=97248481; PubMed=9094709;
RX Perrimon N., Perkins L.A.;
RA "There must be 50 ways to rule the signal: the case of the *Drosophila*
RA EGF receptor";
RL Cell 89:13-16(1997).
CC -1- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOSEROSA
CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
CC CUTICLE.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
CC PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND
CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
CC AND THORACIC AND ABDOMINAL GANGLIA.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AF052754; AAC08536.1; JOINED.
CC EMBL; AF052753; AAC08536.1; JOINED.
CC EMBL; AF052754; AAC08535.1; JOINED.
CC EMBL; AF052752; AAC08535.1; JOINED.
CC EMBL; K03054; AAA51462.1; JOINED.
CC EMBL; K03417; AAA51460.1; JOINED.

DR EMBL; K03416; AAA50965.1; JOINED.
DR EMBL; K03418; AAA51461.1; JOINED.
DR EMBL; AF109077; AAD26134.1; JOINED.
DR EMBL; AF109078; AAD26132.1; JOINED.
DR EMBL; AF109082; AAD26132.1; JOINED.
DR EMBL; AF109078; AAD26133.1; JOINED.
DR EMBL; AF109084; AAD26133.1; JOINED.
DR EMBL; AF109079; AAD26130.1; JOINED.
DR EMBL; AF109081; AAD26130.1; JOINED.
DR EMBL; AF109079; AAD26131.1; JOINED.
DR EMBL; AF109083; AAD26131.1; JOINED.
DR EMBL; AF109080; AAD26135.1; JOINED.
DR EMBL; AF109080; AAD26135.1; JOINED.
DR EMBL; X02293; CAA26157.1; JOINED.
DR EMBL; X78920; CAA55523.1; JOINED.
DR EMBL; X78918; CAA55521.1; JOINED.
DR EMBL; X78919; CAA55522.1; JOINED.
DR PIR; A00640; GQFFE.
DR HSP; P11362; IFGK.
DR FlyBase; FBgn0003731; Egfr.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recept_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
DR PROSITE; PS00109; PROTEIN KINASE_DOM; 1.
DR PROSITE; PS00011; PROTEIN KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
DR Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
DR Developmental protein.
DR SIGNAL 1 30
DR CHAIN 31 1426
DR DOMAIN 31 868
DR TRANSMEM 869 889
DR DOMAIN 890 1426
DR DOMAIN 938 1198
DR NP_BIND 944 952
DR BINDING 971 971
DR ACT_SITE 1063 1063
DR MOD_RES 902 902
DR Query Match 28.0%; Score 1919; DB 1; Length 1426;
DR Best Local Similarity 32.5%; Pred. No. 6.9e-95;
DR Matches 465; Conservative 185; Mismatches 436; Indels 344; Gaps 43;
Qy 24 QVCTGDMKRLRLPASPETHDLRLHLYOGCVVGNLELTYLPT-NASLSPLQIDIOEVQF 82
Db 100 KICIGTSRLSVSPSNKEHYRNLDRVTNCTYVDGNLKLTLWLPNENLDLSPLDNIREV-- 157
Qy 83 NNFTVFWLRVPKVSASHLEQRLIVRGTOLF-----EDNYALAVLNGDPLNNTPTVG 137
Db 158 TGYILISHVDVKVY-----VFPQLIIRGTLFSLSVSEKYLEV----- 198
Qy 138 ASPLGLRELQRLSLTEILKGVLIQNPQLCYQDITLWKDIFHKNNQALTLIDTNRSA 197
Db 199 -TVSKMYTLIPLDLRDVLNGVQVPHNNYNLCHMTIOWSEIVSNGTDAYNYDTAPERE 257
Qy 198 CHPCSPMCKGSRGWSESSDQSLTRTVVAGGCA--RCKGPLPTDCHEQCAAGCTGPKH 255
Db 258 CPKCHESCTHG-CWEGGPKNCKQFSLCTSPQACAGRCYGPCKPRECHLFCAGCTGCTPQ 316
Qy 256 SDCLACLHFNHSGICELHCPALVTYNTDTFESMNPGRYTFGASCYTCAPYNYLSDVVG 315
Db 317 KDCIACKNFFDEAVSKSECPMRKNYPTTYVLETNPEGKYAYGATCVKCEP-GHLLRDNG 375
Qy 316 SCTLVCLPHNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVAVTSANIQBFAGCKK 375

```
Db 376 ACVSCPDQKMDKGE-----CVCNGPCPKTCGVTVLH-----AGNIDSRNCTV 422
Qy 376 IFGLAFIPESFDG--DPASNTA-----PLQPEQLQVETLEEITGYLYISAWPDSLDP 427
Db 423 IDGNIRILDQFSGFDVYANTWGPRIPLDPERREVSFVKETGYLNTIEGTHQPN 482
Qy 428 LSVFQNLQVIRIHNQAY-SLTQLQGLIGISWLGRLSRLREGLSGLALIHNNHLCFVHTV 486
Db 483 LSYFRNLFTIHRQLMESFAALAIVKSSLSLEMRNLKQISSGVSIVTQHNRLDLYVSI 542
Qy 487 PWDOLFNRPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPPTQCVNCSQFLRGEC 546
Db 543 RWPAIQKEPEQKQWVNNRDLCEKNGTICSDQCNEDGCGAGTDCCLTKCNFNFGTC 602
Qy 547 VEECRVLQGLPREYVYNAHCLPCHPEQOPQNGSVTCFGEADQCVACAHYKDPFPCVARC 606
Db 603 IADCGYSNAVK--FDNRCKICHPECR-----TCNGAGADHCQECVHVVDGQHVCSEC 654
Qy 607 P-----SGVK-----PDL 614
Db 655 PKNYNDRGVCRECHATCDGCTGPKDTIGAGCTTCNLAIINNDATVKRCLLKDDKCPD- 713
Qy 615 SYMPIWKF--PDEGACOP-----CPI-----NCTH----- 638
Db 714 GY--FWYVHPQEQGSUKPLAGRAVCRKCHPLCELCTNNGYHQVCSKCTHYRRQOET 771
Qy 639 -----SC-----VLDLDDK----- 647
Db 772 ECPADHYTDEQRBCFQHPCEGCTGPGADDCKSCRNFKLFANETGPGYVNTWFNCTS 831
Qy 648 -CPAEOR-----ASPLTS-----IVSAVWGILLVVVLGVVFGI 679
Db 832 KCPLMRHVNQYTAIGYCAASPPRSKITANLDVNMFIITGAVLVPTICILCV--T 889
Qy 680 LIKROQKIRKYT--MRLLQETELVBLTSGAMPNQAMRILKETELRKVKVLGSGAF 737
Db 890 YICQKQAKKETVMTWALSGCEDSEPLRESNTGANCLKRIYKDAELRGGVGLMGAF 949
Qy 738 GTVYKGIWIPGENVKIPVAIKVLRNTPSKANKEIIDEAAYVAGVSPYVSRLLGCLT 797
Db 950 GRVYKGVVPGENVKIPVAIKELKSTGASSESEFLEAYIMASEHNLLKLLAVCMS 1009
Qy 798 STVOLVTQMPYGCLLDHRVNRGLSGQDILLNCMQIAKMSYLEDVRLVHRDLAARNV 857
Db 1010 SQMMLITQMLPLGCLLDVYRNNDKIGSKALLNWSQIAKMSYLEEKRLVHRDLAARNV 1069
Qy 858 LVKSPNHVKITDFGLARLLDIDETEHADGKVIKWMALSIILRRPHTHQSVDWSVGT 917
Db 1070 LVQTPSLVKITDFGLAKLLSSDSNEYKAAAGKMPKWLALCIRNVFTSKSDYVAFGVT 1129
Qy 918 VWELMTFGKPYDGIIPAREIPDLLEKGERLPPPICTIDVYIMVVKWMIDSECRPRFRE 977
Db 1130 IWELLTQGRPHENIPAKDIPDLLEVGLKEQPEICSDIYCTLLSCLHLDAAWRPTFKQ 1189
Qy 978 LVSEFSRMARDPQRFVVIQNEDLG--PASPLDSTFYRSLLEDD---DMGDLVDVAEEYLV 1032
Db 1190 LTTVFAEFARDPGRYLAIPGKFTLPA-----YTSQDEKDLIRKLAPTTDGSALAK 1242
Qy 1033 QOGFFCDPAPACGMVHRHSRSTSGGDDLTLGLEPSEBAP-----RSLAPSEB 1086
Db 1243 PDDYLQPKAAPGPS-----HRTDCT-----DEMPKLNRYCKDPSNKNSS 1281
Qy 1087 AGSDVFDG---DLGWAAGKGLQSLPHTDPSPLQRYSEDPVPLPSEDTGGVYVAPLTCSPQ 1143
Db 1282 TGDDESSAREVGNLR-----LDLPVDEDDYLMPT-TCQGP 1319
Qy 1144 EYVNPQDVRPQPPREGPLPAARFAGATLERAKTLSPGKNGVVKVDFAFGGAVENPEYL 1203
Db 1320 NNNNNMN-----NPNQNNMAAVGAAGYM-----DLIGVPVSDNPEYL 1358
Qy 1204 ---TPGGGAAPQH-----PPAFSP-AFNNLYYWD 1230
```

Db 1359 LNAQTLGVGSPITQTIGIPVMGGPGTMEVKVPMGPSEPTSSDHEYND 1408

RESULT 12

```
ERBB ALV
ID ERBB ALV STANDARD; PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN V-ERBB.
OS Avian leukosis virus.
OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nilsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA Crittenden L.B., Raines M.A., Kung H.-J.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor.";
RL Cell 41:719-726 (1985).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M10066; AAA48763.1; ALT_INIT.
CC PIR; A00643; TVCHLV.
CC PIR; B00643; TVFVLV.
CC HSP; P11362; IFGK.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00219; Tyrc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.
CC PROSITE; PS00109; PROTEIN KINASE TYR; 1.
CC PROSITE; PS00011; PROTEIN KINASE DOM; 1.
CC Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
CC Glycoprotein; Phosphorylation.
CC DOMAIN 132 399 PROTEIN KINASE.
CC NP BIND 138 146 ATP (BY SIMILARITY).
CC FT BINDING 165 165 ATP (BY SIMILARITY).
CC ACT_SITE 257 257 BY SIMILARITY.
CC SQ SEQUENCE 634 AA; E705E33A0BE01FCC CRC64;
```

Query Match 25.6%; Score 1749.5; DB 1; Length 634;
Best Local Similarity 52.3%; Pred. No. 2.9e-86;
Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17;

Qy 593 CAHKQKPPFCVAPCSGVKFDLSVMPWKPFDEEGACQPCINCHSCVDLDDKGCFAEQ 652

Db 3 CAHFDGPHCVKACPGVLGENDTL-VWKYADANAVCOLCHPNCTRCCKGFGLEGCP--- 58

Qy 653 RASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQKIRKYTMRRLLQETELVPLTPSGA 711


```

Db 59 NGSKTSPSIAAGVGGCLLVVGLGIGLYLRR-HIVKRTLRLLQERLVEPLTPSGE 117
Qy 712 MPNOAQRILKETELRKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANK 771
Db 118 APNOAHLRIKETEFKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANK 177
Qy 772 EILDEAYVMAGVSPYVSRLLGICLTSTVOLVTPMPCGLLDHVRNCRGLSGDILNW 831
Db 178 EILDEAYVMASVDNPHVCRLLGICLTSTVOLVTPMPCGLLDHVRNCRGLSGDILNW 237
Qy 832 CMQIAGKMSYLEDVRLVHRDLAARNVLKSPNPKITDFGLARLLDIDETEHADGGKVP 891
Db 238 CVQIAGKMNYLEERLVRDLAARNVLKTPQHVKITDFGLAKLGADEKEYHAEGGKVP 297
Qy 892 IKWMALESILRRRPTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGRRLPQPP 951
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGPAREIPDLLEKGRRLPQPP 357
Qy 952 ICTIDVYMWKCMIDSECPRELVSERFMRARDPQRFVVIQ-NEDLGPASPLDSTF 1010
Db 358 ICTIDVYMWKCMIDSECPRELVSERFMRARDPQRFVVIQ-NEDLGPASPLDSTF 417
Qy 1011 YRSILLEDDMDGLVDAEYLVPOQGFPCPDPAAGGMVHRRSSSTRSGGDLTLGLE 1070
Db 418 YRTLEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1071 PSEBAPRSP-----APSEGAGSDVFDGLGMAKGLQLSLPHDPSPLQRYSEDPTVP 1125
Db 450 -----SRTPLLSLSLSSATNSATNCID-----RNGQGHVREDSFVQRYSSDPTGN 495
Qy 1126 LPSET-DGYVAPLTCSPQEVNQDVRPQPPSPREGPLPAARAGATLERAKTLPSPK 1183
Db 496 FLESIDGFL-----PAPEYVQ--LMPKKPS-----TAMVQ 526
Qy 1184 NGVVKDVP-----AFGAVENPEYLTPQGAAPQHPPPAPAFDNLYY 1228
Db 527 NQIYNNISLTAISKLPMDSRVQNSHSTAVDNPYL-----NTNQSPKATVFESSPY 578
Qy 1229 WDQ-----DPPE-----RGAPSTFGTPTAENPEYLGIDVP 1260
Db 579 WIQSGNHQINLDNPDYQODFLPNETKPNGLLKVPAEENPEYLRVAAP 625

RESULT 13
ID ERBB_AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN [1]
RC SEQUENCE FROM N.A.
RA MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbb gene of avian erythroblastosis virus is a member of the src
RT gene family."
RL Cell 35:71-78 (1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328658;
RX Debutre B., Henry C., Benaisa M., Biserte G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erba gene of avian erythroblastosis virus reveals a
RT new type of oncogene."
RL Science 224:1456-1459 (1984).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein

```

```

CC tyrosine phosphate.
CC -!- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
CC -!- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; K02006; AAA42394.1; ALT INIT.
CC EMBL; K01216; AAA42400.1; -.
CC
CC PIR; A0644; TVYUHK.
CC HSP; P11362; IFGK.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.
CC PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
CC PROSITE; PS00111; PROTEIN KINASE_DOM; 1.
CC Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 146 146 I -> V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MW; 76EBDD06745D609 CRC64;

Query Match 24.9%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 8.2e-84;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

Qy 593 CAHYKDPFPFCVAPCGVGPDLSPYPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQ 652
Db 3 CAHFDGPHCVKACPAVGLGENDTL-VRKYADANAVCQLCHPNCRTCKGPGLEGCP--- 58
Qy 653 RASPLTSIVSAVV-GIILVVVLGVVFGILIKRRQKIRKYMRLLOETELVEPLTPSGA 711
Db 59 NGSKTSPSIAAGVGGCLLVVGLGIGLYLRR-HIVKRTLRLLQERLVEPLTPSGE 117
Qy 712 MPNOAQRILKETELRKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANK 771
Db 118 APNOAHLRIKETEFKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANK 177
Qy 772 EILDEAYVMAGVSPYVSRLLGICLTSTVOLVTPMPCGLLDHVRNCRGLSGDILNW 831
Db 178 EILDEAYVMASVDNPHVCRLLGICLTSTVOLVTPMPCGLLDHVRNCRGLSGDILNW 237
Qy 832 CMQIAGKMSYLEDVRLVHRDLAARNVLKSPNPKITDFGLARLLDIDETEHADGGKVP 891
Db 238 CVQIAGKMNYLEERLVRDLAARNVLKTPQHVKITDFGLAKLGADEKEYHAEGGKVP 297
Qy 892 IKWMALESILRRRPTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGRRLPQPP 951
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGPAREIPDLLEKGRRLPQPP 357
Qy 952 ICTIDVYMWKCMIDSECPRELVSERFMRARDPQRFVVIQ-NEDLGPASPLDSTF 1010
Db 358 ICTIDVYMWKCMIDSECPRELVSERFMRARDPQRFVVIQ-NEDLGPASPLDSTF 417
Qy 1011 YRSILLEDDMDGLVDAEYLVPOQGFPCPDPAAGGMVHRRSSSTRSGGDLTLGLE 1070

```



```

Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1071 PSEEARPSPL-----APSEGAGSDVFDGLMGAAKGLQLPHTDPSPLQRYSEDPTVP 1125
Db 450 -----SRTPLLSLSLTSATSNNSATNCID-----RNGQGHVREDSEFVQRYSSDPTGN 495
Qy 1126 LPSET--DCYVAPLTCSPQPEYVQDPVVRPOPPSPREGPLPAARPAGATLSEKTLSPGK 1183
Db 496 FLEESIDGFL-----PAPEYVQ--LMPKPKSTAM----- 524
Qy 1184 NGVVKDVFAP-----GGAVENPEYLTPOGGAAPQHPHPPAFSPAFD 1224
Db 525 --VQNOIYNFSLTAISLKLPMDSRYQNSHSTAVDNPEYL-----NTNOSPLAKTVFE 574
Qy 1225 NLYYWDQPPPERGAPPSTFKGTPTAENPEY 1254
Db 575 SSPYIQSGNHQ-----INLDNPDY 594

RESULT 14
ID ERBB AVIEU STANDARD; PRT; 540 AA.
AC P11273;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ts167).
OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=103898;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87064458; PubMed=2878364;
RA Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;
RT "A single amino acid substitution in v-erbB confers a thermolabile
RT phenotype to ts167 avian erythroblastosis virus-transformed erythroid
RT cells."
RL Mol. Cell. Biol. 6:1751-1759(1986).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M13179; AAA42401.1; -.
CC PIR; A25231; TVFVEB.
CC HSP; P11362; IFGK.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
CC Glycoprotein; Phosphorylation.
CC DOMAIN 132 399 PROTEIN KINASE.
CC FT NP_BIND 138 146 ATP (BY SIMILARITY).
CC FT BINDING 165 165 ATP (BY SIMILARITY).
CC FT ACT_SITE 257 257 BY SIMILARITY.
CC FT VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB).
CC SQ SEQUENCE 540 AA; 60412 MW; 5B53297AA068B65D CRC64;

Query Match 23.8%; Score 1630; DB 1; Length 540;
Best Local Similarity 55.1%; Pred. No. 5.6e-80;

```

```

Matches 341; Conservative 69; Mismatches 119; Indels 90; Gaps 14;
Qy 593 CAHYKDPFPCVACPSGVKFDLSYMPIWKFPDEBGAQCPQPCINCHSCVDLDDKGCRAEQ 652
Db 3 CAHFIDGPHCVKACPAVGLGENDTL-VWKYADANAVCOLCHPNCNTRGCKGPGLEGCP--- 58
Qy 653 RASPLTSIVANV--GILLVVLGVVFGVGLIKRROOKIRKYTMRLLOETELVELPLTSGA 711
Db 59 NGSKTPTSAAGVGLGCLLVVVGIGLGLYLR--HIVKRTLRLLQERELVELPLTSGE 117
Qy 712 MPNOAQRILKETELRKVKVLGSGAGCTVYKGIWIPGENVKIPVAIKVIRENTSPKANK 771
Db 118 APNOAHURILKETEFKVKVLGFGAGCTVYKGLWIPGEKVTIPAIKELREATSPKANK 177
Qy 772 EILDEAYVMAGVSPYVSRLLIGICLTSTVOLVTLMPYGCLLDHRVNRGLGSDLLNW 831
Db 178 EILDEAYVMASVDNPHVCRLLIGICLTSTVOLITQLMPYGCLLDYIREHKONIGSQYLLNW 237
Qy 832 CMQIAKMSYLEVDVRLVHRDLAARNVLKSPNVHKITDFGLARLLDIDETVHADGGKVP 891
Db 238 CVQIAKGMNYLEERHVMVHRDLAARNVLKTPQHVKITDFGLAKQLGADEKHYAEGGKVP 297
Qy 892 IKMALESILFERETHOSDVMSYGVTVWELMTFGAKPYDGTIPAREIPDLLEKGERLP 951
Db 298 IKMALESILERIYTHOSDVMSYGVTVWELMTFGSKPYDGTIPASEISSVLEKGERLP 357
Qy 952 ICTIDVYMWKCMIDSECRPRFRELVSFSEMRARDPQRFVVIQ-NEDLGPASPLDSTF 1010
Db 358 ICTIDVYMWKCMWSADSRPKRELIAEFKWARDPPRYLVLIQGERMHLPSFTDSKF 417
Qy 1011 YRSLLDDDDGLVDABEYLVPOGPFCDPAPAGAGMGVHHRSSSTRSGGDLTLGLE 1070
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1071 PSEEARPSPL-----APSEGAGSDVFDGLMGAAKGLQLPHTDPSPLQRYSEDPTVP 1125
Db 450 -----SRTPLLSLSLTSATSNNSATNCIDRNG-----H----- 476
Qy 1126 LPSETDGVAPLTCSPQPEYVQDPVVRPOPPSPREGPLPAARPAGAT--LERAKTSLPGKN 1184
Db 477 -PVREDGFL-----PAPEYVQ--LMPKPKSTAMVQNIYVSLTAISKLPIDSRYQN 527
Qy 1185 GVVKDVFAGGAVENPEYL 1203
Db 528 -----SHSTAVDNPEYL 539

RESULT 15
ID EGFR CHICK STANDARD; PRT; 703 AA.
AC P13387;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER
DE (Fragment)).
GN EGFR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88261272; PubMed=3260329;
RA Lax I., Johnson A., Hawk R., Sap J., Bellot F., Winkler M.,
RA Ullrich A., Vennstrom B., Schlessinger J., Givol D.;
RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT expression in mouse cells, and differential binding of EGF and
RT transforming growth factor alpha."
RL Mol. Cell. Biol. 8:1570-1578(1988).
CC -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND

```

CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -|- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC -|- TYROSINE PHOSPHATE.
CC -|- SUBCELLULAR LOCATION: Type I membrane protein.
CC -|- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -|- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M20386; AAA48760.1; ..
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; FU; 4.
DR PROSITE; PS00107; PROTEIN KINASE ATP; PARTIAL.
DR PROSITE; PS00109; PROTEIN KINASE TYR; PARTIAL.
DR PROSITE; PS50011; PROTEIN KINASE DOM; PARTIAL.
KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 30
FT CHAIN 31 >703 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT DOMAIN 31 654 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 655 667 POTENTIAL.
FT DOMAIN 668 >703 CYTOPLASMIC (POTENTIAL).
FT DISULFID 197 206 BY SIMILARITY.
FT DISULFID 201 214 BY SIMILARITY.
FT DISULFID 222 230 BY SIMILARITY.
FT DISULFID 226 238 BY SIMILARITY.
FT DISULFID 239 247 BY SIMILARITY.
FT DISULFID 243 255 BY SIMILARITY.
FT DISULFID 258 267 BY SIMILARITY.
FT DISULFID 271 298 BY SIMILARITY.
FT DISULFID 302 314 BY SIMILARITY.
FT DISULFID 318 333 BY SIMILARITY.
FT DISULFID 336 340 BY SIMILARITY.
FT DISULFID 513 522 BY SIMILARITY.
FT DISULFID 517 530 BY SIMILARITY.
FT DISULFID 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT DISULFID 565 581 BY SIMILARITY.
FT DISULFID 569 589 BY SIMILARITY.
FT DISULFID 592 601 BY SIMILARITY.
FT DISULFID 605 627 BY SIMILARITY.
FT DISULFID 630 638 BY SIMILARITY.
FT DISULFID 634 646 BY SIMILARITY.
FT CARBOHYD 134 134 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 190 190 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 359 359 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 420 420 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 613 613 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 633 633 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 648 648 N-LINKED (GLCNAC. .) (POTENTIAL).
FT NON_TER 703 703
SQ SEQUENCE 703 AA; 77427 MW; AFF2DE11B735A690 CRC64;
Query Match 22.7%; Score 1555; DB 1; Length 703;
Best Local Similarity 43.3%; Pred. No. 7, 5e-76;

Matches 309; Conservative 115; Mismatches 255; Indels 34; Gaps 15;
Qy 8 RWGLLLALLPPGAA-----STQVCCTGDMKRLRASPETHLDMRLHYQGCQVVQGNLE 61
Db 13 RGAALVLLLLGVALCSAVEEKVCQGTNNKLTQLGHVEDHFTSLQRMNVCNEVLSNLE 72
Qy 62 LTYLPTNASLFLQDIOEVQFNFTVSFWLRVPKVSASHLEQRLRIVRGTFQLEFDNYALA 121
Db 73 ITVEHNRDITFLKTIOEVA--GYVL---TALNVVDVIPLE-NLQIIRGNVLYDNPALA 126
Qy 122 VLNGDPLNNTPTVGTASPGGLRELQRLSITELKGVLIQRLNPOLCYQDTILMKDIFHK 181
Db 127 VLSNYH-MNKTQ-----GLRELPMKRLSEILNGVKISNNPKLNCMDTIVLWMDIIDT 177
Qy 182 NNQLATLID-TNRSRACHPCSPMKSGRCWGESSEDCOSLTRTVCAAGGA-RCKGPPLPT 239
Db 178 SRK-PLTVLDFASNLSSCPKCHNCTEDHCWAGBQNCQTLTKVICAQCSGRCRGKVP 236
Qy 240 DCHEQCAAGCTGPKHSDCLACLHFNHSGICEHCPALVTYNTDTFESMPNPEGRYTFGA 299
Db 237 DCHNQCAAGCTGPRESDCLACKFRDDATCKOTCPPLVLYNPTYQMDVNPCKYSEGA 296
Qy 300 SCVTACPYNYLSTVDGSCITLVCPLHNOEVTAEQTCCEKSKPCARVCYGLGMEHREV 359
Db 297 TCVRCEPHNVVVDHSGSVRSCTNTDYEV--EENGVRKCKKCDGLCSKVCNGIGIGELKGI 355
Qy 360 RAVTSANIQEFAGCKIFGSLAFLPESFDGDPASNTAPLOPELOQVFETLEEITGYLYIS 419
Db 356 LSNATNIDSFKNCTKINGDVSLPVAFLGDFTKTLPLDPKLDVFTVKESIGFLLIQ 415
Qy 420 AWPDSLPLDSVFQNLQVIRGRIHLHGAYSILTLQGLGISWLGRLSLRELGSGLALIHNT 479
Db 416 AWPDNATDLYAFENLEIIRGTRKHGOYSLAVNVLKIOSLGLRSLKEISDGDIAIMKN 475
Qy 480 LCFVHTVPWDOLFRNPHQALLHTANRPEDECVGEGLAACHOLCARGHCWGPGPTQVCNCSQ 539
Db 476 LCVADTMNRSLFATQSKTKIQRNRKNDCTADRHVCDPLCSDVGCWGPFPFCFSCRF 535
Qy 540 FLRQCECVESCRVLOGLPREYVNAHCLPCHPECPONG---SVTCFGEADQCAHAHY 596
Db 536 FSRQKECVKQCNILQGEPRFERDSKCLPCHSECLVQNSTAYNTTCSGPGPDHCKMCAHF 595
Qy 597 KDPFPCVARCPSGVKPDLSYMPIWKFPDEBACQPCPCINCTHSCVDLDDKCPAEQRASP 656
Db 596 IDGPHCVKACPAAGVLGENDTL-VWKYADANAVCOLCHPNCNTRGCKGPGLEGCP---NGSK 651
Qy 657 LTSIVSAVW-GILLVVVLGVVFGILIKRQOKIRKYMRLRLLOTELVEPLTP 708
Db 652 TPSIAGWVGGLCLVVGVLGIGLYLRRR-HIVRKRTLRLRLQRELVEPLTP 703

Search completed: July 22, 2003, 08:44:14
Job time : 21.2793 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.5887 Seconds
(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-59-73-14
Perfect score: 6847
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGGLDVPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6202	90.6	1259	6	O18735
2	3101	45.3	1209	11	O9QX70
3	2072	44.9	1210	11	O9EP98
4	2729.5	39.9	1137	13	O9W6F6
5	2683	39.2	1165	13	O9YH40
6	2263	33.1	1328	13	P79754
7	1959.5	29.2	1433	5	O9BIH9
8	1775.5	25.9	419	4	O9UK79
9	1739	25.4	367	11	O8RX1
10	1720	25.1	729	15	O86712
11	1718	25.1	567	15	O86714
12	1697.5	24.8	412	4	O8WYV0
13	1653.5	24.1	962	15	O64895
14	1645	24.0	545	15	O85468
15	1463.5	21.4	655	11	O9WVF5
16	1447.5	21.1	643	11	O9ERV6

17	1269	18.5	1193	5	O9YIX8
18	1172.5	17.1	1368	5	O23821
19	1121	16.4	1717	5	O26566
20	1086	15.9	527	13	O90836
21	973.5	14.2	478	11	O9SE0
22	906	13.2	165	4	Q14256
23	902.5	13.2	599	13	O9F5H2
24	887	13.0	176	11	O923V5
25	806.5	11.8	346	13	P11776
26	778	11.4	435	5	O8SZW1
27	757.5	11.1	1362	13	O9PVZ4
28	754.5	11.0	311	13	O99162
29	732	10.7	1671	5	O9NVV5
30	723	10.6	149	6	O9BG66
31	712	10.4	1418	13	O93457
32	706	10.3	1358	13	O73798
33	702.5	10.3	1368	13	O8UW85
34	678.5	9.9	1369	13	O8UW86
35	676	9.9	331	4	O9BUD7
36	668	9.8	1412	13	O8UW84
37	667.5	9.7	1418	13	O8UW83
38	658	9.6	1371	11	O9QVW4
39	645	9.4	1245	13	O9YGH8
40	643.5	9.4	1472	5	O9USA8
41	617.5	9.0	2144	5	O9VD94
42	604	8.8	987	11	O91YM0
43	603	8.8	935	4	O96L35
44	601	8.8	987	11	O99WR2
45	587.5	8.6	1036	4	Q07912

ALIGNMENTS

RESULT 1

O18735	O18735	PRELIMINARY;	PRT; 1259 AA.
AC	O18735;		
DT	01-JAN-1998 (TrEMBLrel. 05, Created)		
DT	01-JAN-1998 (TrEMBLrel. 05, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	ErbB-2.		
OS	Canis familiaris (Dog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.		
OX	NCBI_TaxID=9615;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Yokota H.;		
RT	"CDNA cloning of erbB-2 from canine mammary gland.";		
RL	Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AB008451; BAA23127.1;		
DR	HSSP; P11362; IFGK.		
DR	InterPro; IPR002048; EF-hand.		
DR	InterPro; IPR000494; EGFR_L_domain.		
DR	InterPro; IPR000719; Euk_pkinase.		
DR	InterPro; IPR002174; Furin-like.		
DR	InterPro; IPR001245; Tyr_pkinase.		
DR	InterPro; IPR004019; YLP_motif.		
DR	Pfam; PF00757; Furin-like; 1.		
DR	Pfam; PF00069; pkinase; 1.		
DR	Pfam; PF01030; Recep_L_domain; 2.		
DR	Pfam; PF02757; YLP; 3.		
DR	ProDom; PD000001; Euk_pkinase; 1.		
DR	SMART; SM00261; FU; 3.		
DR	SMART; SMO0219; TyrKc; 1.		
DR	PROSITE; PS00018; EF_HAND; UNKNOWN 1.		
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.		
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.		
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.		
KW	ATP-binding; Transferase; Tyrosine-protein kinase.		
SQ	SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;		

Query Match		90.6%; Score 6202; DB 6; Length 1259;
Best Local Similarity		90.7%; Pred. No. 0;
Matches 1148; Conservative 42; Mismatches 64; Indels 12; Gaps 4;		
Qy	1	MELAAALRWGGLLLALLPPGAASVCTGTDMLKRLPASPETHDMLRLHYGCGVQVGNL 60
Db	1	MELAAALRWGGLLLALLPPGAAGTCTGTDMLKRLPASPETHDMLRLHYGCGVQVGNL 60
Qy	61	ELTYLPTNASLFTODIOEVQFNFTVSFWLRVKVSASHLEORLIRVRGTLQDFEDNAL 120
Db	61	ELTYLPANASLFLQDIOEVQ--GYLTAHSQVRQIPL---QRLIRVRGTLQDFEDNAL 114
Qy	121	AVLNDGDPNNTPVTGASPGGLRELQRLSITEILKGGVLIQRPOLCYQDITLWKDIFH 180
Db	115	AVLNDGDPLEGGIPAPGAHQOGLRELQRLSITEILKGGVLIQRPOLCHQDITLWKDVFH 174
Qy	181	KNNQALFLIDTNRASRACHPCSPCKGSRCKGSESSDCQSLTRTVCAAGGCARCKGPLPTD 240
Db	175	KNNQALFLIDTNRFSACPPCSPACKDAHCWGASSGDCQSLTRTVCAAGGCARCKGPOPTD 234
Qy	241	CCHQCAAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESHPNPEGRYTFGAS 300
Db	235	CCHQCAAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESHPNPEGRYTFGAS 294
Qy	301	CVTACPNYLTDSVGSCTLVCPHNOEYVTAEDGTQRCCKSPCARVCYGLGMEHLREVR 360
Db	295	CVTSCPNYLTDSVGSCTLVCPHNOEYVTAEDGTQRCCKSPCARVCYGLGMEHLREVR 354
Qy	361	AVTSANIOEAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFTLEBITGYLYISA 420
Db	355	AVTSANIOEAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLRVFEALEBITGYLYISA 414
Qy	421	WPSLPLSLVFQNLQVIRGRILHNGAYSITLQGLISWLGRLSRELGSGLALIHNNTHL 480
Db	415	WPSLPLSLVFQNLQVIRGRILHNGAYSITLQGLISWLGRLSRELGSGLALIHNRAL 474
Qy	481	CFVHTVPMDQLFRNPQOALLHTANRPECECVGEGGLACHOLCARGHCWGPQTQVNCQSF 540
Db	475	CFVHTVPMDQLFRNPQOALLHSANRPECECVGEGGLACP--CAHGCWGFQTQVNCQSF 533
Qy	541	LRGQCEVBEERVLQGLPREYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDDP 600
Db	534	LRGQCEVBEERVLQGLPREYKDYCLPCHSECPQNGSVTCFGEADQCVACAHYKDDP 593
Qy	601	FCVARCPGVKPDLSYMPIWFPDEBEGACQCPINCHTSCVDLDKGCAPQORASPLTSI 660
Db	594	FCVARCPGVKPDLSFMPWIWFADEBEGTCPCPINCNTSCADLDEKGCAPQORASPLTSI 653
Qy	661	VSADVGLLVVGVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAOMRI 720
Db	654	IAAVVGILLAVVGLVGLILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAOMRI 713
Qy	721	LKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVIRENTSPRANKEILLDEAYVM 780
Db	714	LKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVIRENTSPRANKEILLDEAYVM 773
Qy	781	AGVGSPPVSRLLGICLTSTVQLMPYGLLDHVRNRRGLSQDILLNCKMIAKMS 840
Db	774	AGVGSPPVSRLLGICLTSTVQLMPYGLLDHVRHRRGLSQDILLNCKVQIAKMS 833
Qy	841	YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKKNWALESI 900
Db	834	YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKKNWALESI 893
Qy	901	LRRRFTHQSDVMSYGVTVWELMTGAKPYDGPAREIPDLLEKGERLPQPPICITDVYMI 960
Db	894	PRRFTHQSDVMSYGVTVWELMTGAKPYDGPAREIPDLLEKGERLPQPPICITDVYMI 953
Qy	961	MVKCMWIDSECRPRFRELVSFESRMARDPQRFVIONEDLPASPLDSTFYRSLLEDDEM 1020
Db	954	MVKCMWIDSECRPRFRELVAEFSRMARDPQRFVIONEDLPASPLDSTFYRSLLEDDEM 1013
Qy	1021	GDVDAEYLVPOQGFCCPDPAPGAGGMVHRRSSSTRSGGDLITLGLPSEBEPKSP 1080
Db	1014	GDVDAEYLVPOQGFCCPPTGAGCTAHRHRSSTRNGGELTGLPSEBEPKSP 1073
Qy	1081	LAPSEGAGSVFDGDLGMAAKGLQSLPHTDPPQLQRYSEDPTVPLPSETDGYVAPLTCS 1140
Db	1074	LAPSEGAGSVFDGDLGMAAKGLQSLPQRYSEDPTVPLPSETDGYVAPLTCS 1133
Qy	1141	POEYVNOQDVRRQPPSPREGPLPAARPAGATLER----AKTILSPCKNGVVKDVAFGG 1195
Db	1134	POEYVNOQDVRRQPPSPREGPLPAARPAGATLERPKTILSPCKNGVVKDVAFGG 1193
Qy	1196	AVENPEYLTQGGAAPOHPHPPAFSPAFDNLVWDQDPPRGAPPSTFKGTPTAENPEYL 1255
Db	1194	AVENPEYLARGRAAPQHPHPPAFSPAFDNLVWDQDPPRGAPPSTFKGTPTAENPEYL 1253
Qy	1256	GLDVPV 1261
Db	1254	GLDVPV 1259
RESULT 2		
ID	Q9QX70	PRELIMINARY; PRT; 1209 AA.
AC	Q9QX70;	
DT	01-MAY-2000 (TrEMBLrel. 13, Created)	
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)	
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)	
DE	Epidermal growth factor receptor.	
GN	EGFR.	
OS	Rattus norvegicus (Rat).	
OC	Eukaryota; Chordata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OX	NCBI_TaxID=10116;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
RC	MEDLINE=90258888; PubMed=2342466;	
RA	Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,	
RA	Earp H.S.;	
RT	"A truncated, secreted form of the epidermal growth factor receptor is	
RT	encoded by an alternatively spliced transcript in normal rat tissue."	
RL	Mol. Cell. Biol. 10:2973-2982(1990).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
RA	Petch L.A.;	
RL	Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
RA	Guttridge K., Dawson T.L., Earp H.S.;	
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; M37394; AAF14008.1; -.	
DR	HSP; P11362; IFGK.	
DR	InterPro; IPR000494; EGFR_L domain.	
DR	InterPro; IPR000719; Euk_pkinase.	
DR	InterPro; IPR002174; Furin-like.	
DR	Pfam; PF00757; Furin-like; 1.	
DR	Pfam; PF00069; pkinase; 1.	
DR	Pfam; PF01030; Recep_L domain; 2.	
DR	PRINTS; PR00109; TYRKINASE.	
DR	ProDom; PD000001; Euk_pkinase; 1.	
DR	SMART; SM00261; FU; 3.	
DR	SMART; SM00219; TyrcK; 1.	
DR	PROSITE; PS00107; PROTEIN KINASE ATP; 1.	
DR	PROSITE; PS00011; PROTEIN KINASE DOM; 1.	
DR	PROSITE; PS00109; PROTEIN KINASE_TYR; 1.	
KW	ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.	
SQ	SEQUENCE 1209 AA; 134891 MW; 96FEF7F6CC1B7773 CRC64;	
Query Match		45.3%; Score 3101; DB 11; Length 1209;
Best Local Similarity		49.4%; Pred. No. 5.9e-226;

DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
KW ATP-binding; Receptor; Transferase.
SQ SEQUENCE 1210 AA; 134840 MW; 62CD021C9D3E2E18 CRC64;

Query Match 44.9%; Score 3072; DB 11; Length 1210;
Best Local Similarity 48.9%; Pred. No. 9.3e-224;
Matches 625; Conservative 171; Mismatches 365; Indels 118; Gaps 25;

QY 11 LLLALLPGAA--STQVCTGTDMLRLPASPETHLDMLRHLYQCQCVVQGNLEITYPTN 68
DB 14 LLLTALCAAGGALEKKVKCGTSNRLTQGTFFDHLFSLQRYNNCEVVLGNLEITYVQRN 73
QY 69 ASLSFLODIOEVO--FNNFTVSFWLRVPKVSASHLEQRIRIVRGQQLPFDNEDYVALAVLNGD 127
DB 74 YDLSFLKTIQEVAGYVLIATNVERIPL-----ENLQIRGNALYENTYVALAILSN-- 124
QY 128 PLNNTTPVTGASPGGLRELQRLSLEILKGVLIQRPOLCYQDTILWKDI---FHKN 183
DB 125 -----YGNRTGLRELPMRNQELIGAVRFSNNPILCNMDTIQWRDIVQVFNW 176
QY 184 QLALTLIDNRSRACHPCSPMKGSRGWSESDCSQSLTRTVACAGCA--RCKGPLPTDCC 242
DB 177 SMDL-----QSHPSPCPKDPCSPNGSCWGGGEENCQKLTIIICAQCSHRCGRSPSDCC 232
QY 243 HEQCAAGCTGPKHSDIACLHFNHSGICELHCPALVTYNTDTPFSMNPEGRYTFGASCV 302
DB 233 HNQCAAGCTGPRSDCLVQKQFDEATCKTCTPMLLYNPTTYQMDVNPGEKYSFGATCV 292
QY 303 TACPNYLSLDVGCTLVCPLNHOEVAEDTQCEKSKPCARVCYGLGMEHLREVRVAV 362
DB 293 KKCPRNVVYTHGSCVACRAGDPDYEV--EEDGIRKCKDGPCKVCGNGIGEGEFKDTLSI 351
QY 363 TSANIQFAGCKTIFGSLAFIPESFDGDPASNTAPLOPEQIQVFTLEETIGYLIYISAWP 422
DB 352 NATNKHFKYCTAISGLHILPVAFKGSFRTPLDLPRELEILKTVKEITGFLIIQAMP 411
QY 423 DSLPDLVFQNLQVIRGLHNGAYSILTQGLGTSWGLRSLRELGLALIHNTHLCHF 482
DB 412 DNWTDLHAFENLEIRTKQHQGFSLAVVGLNITSLGRLSLEISDGDVLIISGNRLCY 471
QY 483 VHTVPWDLRPNHQAULLHTANREDEECVGLGLACHOLCARGHCWGCPCTOCVNCQFLR 542
DB 472 ANTINWKLFTGNQTKIMNRAEKCKAVHNVNPLCSEGCWGPBPRDCVSCQVSR 531
QY 543 GQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCAVCAHYKDPFPC 602
DB 532 GRECEKCNILLEGPREFVENSEICIQHPECLPOAMNITCTGRGPDNCIQCAHYIDGPHC 591
QY 603 VARCPGVKPLSYMPIWKFPDEGACQCPINCTHSCVDLDDKGPABQASPLTSIVS 662
DB 592 VKTCPAGIMGNNTL--VMKYADANNVCHLCHANCTYGCAGPGLQCEVWPSPGKIPSTAT 650
QY 663 AVUGILLVVLGVVFGI--LIERQOKRKTYMRLLQETELVEPLTPSGAMPNQAQRIL 721
DB 651 GIVGGLLFIVV--VALGIGLFMRHRIVNRKRLRLLQERLEVEPLTPSGEAPNQAHLRIL 709
QY 722 KETELRKVKVLGSGAFGVYKGIWPDGENVKIPVAIKVLRENTSPRANKILDEAYMA 781
DB 710 KETEFKKIKVLGSGAFGVYKGLIPEGEKVIKVAIKELREATSPRANKILDEAYMA 769
QY 782 GVGSPYSRLIGICLTSTVQLVTLQMPYGCILLDHVRENRLGSGDILLNWCQIAKMSY 841
DB 770 SVDNPHVCRLLIGICLTSTVQLITQMPYGCILLDYVREHKDNGISQYLLNWCQIAKGMNY 829
QY 842 LEDVRLVHRDLAARNLVKSPNHVKITDPLGLARLLDIDETVHADGGKVPKKNWALSIL 901
DB 830 LEDRRLVHRDLAARNLVKTPQHKITDPLGLAKLGAEEKYHAEKGVKPKKNWALSIL 889
QY 902 RRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIM 961

890 HRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASDISSILEKGERLPQPPICTIDVYIM 949
962 VKCWMIDSECRPFRELVSFEFSRMARDPQRFVVIQ--NEDLGPASPLDSTFYRSLLEDDDM 1020
950 VKCWMIDADSRPKRELILFEFSKMARDPQRYLVIOGDERMHLPSPTDSNFYRALMDEDM 1009
1021 GDLVDAEYLVPOQGFPPDPAPFAGAGGVHHRHRSSTRSGGDDLTLGLEPSEEA PRSP 1080
1010 EDVVDADEVLTPOQGF-----NSPST-----SRTPT 1035
1081 LAPSEGAGSVDFDGLGMGAAGLQSLPHTDPSFLQRYSEDPTVPLPSET--DGYVAPLT 1138
1036 LLSLSLATS-----NSTVACINRNGSCRVKEDAFLOQRYSSDPTCAVTEIDIDDAFL---- 1087
1139 CSQPEVNVQPDYRPPPPSPREGPLPAARPAAGATLERAKTLSPKNGVVKDVFAFGAVE 1198
1088 --VPPEVNVQ--SVPKRPAGSVQNPVYHNOPLHP-----APGRDLHYQN--PHSNAV 1134
1199 NPEYL--TPQGAAPQPHPPPAFPADNLYWDO-----DP-----PERGAPPS 1241
1135 NPEYLTAQ-----PTCLSSGFSNPAWIKQSHQMSLDNDPDYQODFPFKETKPNG 1185
1242 TFKGTPTAENPEYLGIDVP 1260
1186 IFKG-PTAENAEYLRVAPP 1203

PRELIMINARY; PRT; 1137 AA.

Q9W6F6
AC Q9W6F6;
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DE 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Receptor tyrosine kinase (fragment).
GN ERB4.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HINDERAIN;
RX MEDLINE=99263203; PubMed=10328884;
FA Dixon M., Lumsden A.;
RT "Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in embryonic chick hindbrain."
RL Mol. Cell. Neurosci. 13:237-258 (1999).
DR EMBL; AF121963; AAD31764.1; -.
DR HSP; P11362; IFGK.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001368; TNFR_c6.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 1.
DR PRINTS; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrcK; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
DR PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.
KW Kinase; Tyrosine-protein kinase.
FT NON TER 1
SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC84F CRC64;

```
Query Match 39.9%; Score 2729.5; DB 13; Length 1137;
Best Local Similarity 47.3%; Pred. No. 7.9e-198;
Matches 539; Conservative 169; Mismatches 354; Indels 77; Gaps 24;

Qy 167 LCYODTILWIDIEHKNNQALTLTDNRSRACHPCSPCKGSRGWGESSEDCQSILTRVC 226
Db 3 LCFADTIHWQDIVRNWASNFITLPTNGSSGGRCHKSCCTG-RCWGPFTENHCQILTKTVC 61

Qy 227 AGGC-ARCKGPLPTDCCHQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTYNTDF 285
Db 62 AEQCDGRCYGPVSDCHRECAGCGSKPDKTDFACMNFNDGACVTCQPTQFVFNPTTF 121

Qy 286 ESMNPGRYTFGACVTCACPNYLSTDVSGCTLVCPLHNOBVTAEQTCRCCKSRPCA 345
Db 122 QLEHNHNAKITYGAFCKVKCNHNFV-VDSSSCRACPCSSKMEV-EENGKCKCKCTDTCF 179

Qy 346 RVCYGLGMEHLREVRVTSANIOFAGCKIFGSLAFLPESFDGDPASNTAPLOEQLOV 405
Db 180 KACDGIGTGLSVSAQTVDSSNIDFEINCKTCKINGNLIFLVGIGHGDPHYTIAINPERKNI 239

Qy 406 FETLEEITGYLYISAWPDSLDLSVFQNLQVIRGIRLHNGAYSLTQLGLGTSWLGRLSLR 465
Db 240 FOTVREITGYLNIQSPENMTDFVSNLVTIGGRALYSGLSLLLKQGGITSLFQFQSLK 299

Qy 466 ELGSLALIHNTLHCFVHTVPWQOLFRNPHQALLHTANRPEDCEVGEGLACHQLCARGH 525
Db 300 QISAGNIYITNSNLCTYHTWTSLSFSTPSQKTVIHRNKAENCTADGMVNCNELCSDG 359

Qy 526 CWGPGPTQCVNCSOFLRGQECVSCRVLQGLPREVNAHGLCPHCEQCP-QNGSVTCFG 584
Db 360 CWGPGPOCLCKKFIKRTCTIESCNLYDGFREFANGSVCEWQPCQCKMEDNMNITCYG 419

Qy 585 PEADQCVACAHYKPPFCFVARCPGKVPDLSPYMIWKFPEDEGACQCPINCTHSCVDLD 644
Db 420 PGPBHTKCFHFKDGPNCEKCPDGLQGANSF--IFKYADEDRCHPCPNCTQCGRPA 477

Qy 645 DKGC-----PABQRASPLTISIVAVV-GILLVVLGVVFGILIKRQOKIRKYT 692
Db 478 SHDCIYPTWTRQSTLPQHAR-TPL--IAAGVIGLFIIVIMGLTFAVYVRRKSIK-KKRA 533

Qy 693 MRLLQETELVEPLTPSGAMPNQOMRILKETELRKVKVLGSGAGTGYKGIWTPDGENV 752
Db 534 LRREL-ETELVEPLTPSGATPNQAOQLRKETELRKRVKVLGSGAGTGYKGIWPEGETV 592

Qy 753 KIPVAIKVLRNTPSPKANKEILBAYVMAGVGSFVSRLLIGLITSTVQLVTQLMPYGCIL 812
Db 593 KIPVAIKILNETTGPKANVEFDEALINASMDHPELVLLGVLCSPTLIQVTLMPHGCL 652

Qy 813 LDHVRNRLGSGQDLNWCQIAKMSYLEDEVRLVHRDLAARNVVKSPNHVKITDFGL 872
Db 653 LDYVHEKDNITGSQLLNWCQIAKGMVLEERLVRDLAARNVVKSPNHVKITDFGL 712

Qy 873 ARLLDIDETEYHAGGKVPKWMALLESILRRFTQSDVWSYGVTVWELMTFGAKPYDGI 932
Db 713 ARLLGDEKEYNAGGKVPKWMALLECIFYKFTQSDVWSYGVTVWELMTFGKPYDGI 772

Qy 933 PARBIPDLLEKGERLPQPICTIDVYIMVWCWIMDSECRPRFRELSEFSRMARDPQF 992
Db 773 PTRBIPDLLEKGERLPQPICTIDVYIMVWCWIMDADSRKFELAAEFSEFMARDPQY 832

Qy 993 VVIQNE-LGPASPLDSTFYFSLLEDDMDGLVDAEYLVPOQGFCDPAPGAGGMVHH 1051
Db 833 LVIQDDRMKLPSPNDSKFFQNLLDEEDLMDWDAEYLVF-QAFNIPPIYTSRTRIDS 891

Qy 1052 RHRSSSTRSGGDLTLGLFEPSEERAPS--PLAP-SEGAGSDVFDGLGMAAGKGLQSLP 1108
Db 892 NRNFVYRDGGYAAEQGV-PMPYRAGCIIIEAPVAQAGATAIEFDTCNCTLRKQVATL 950

Qy 1109 THDPSPLQRYSEDPVPLPS-----ETDGYVAPLTCSPQEVYNQDVRPQPPSPREG 1161
Db 951 AKEDSSQRYASDTPVIFPVRVIRGELDEDEGYMTFMRDKPTDYLNVPEENPFVSRKNG 1010

Qy 1162 PLPAA-RPAGATLERAKTSLSPKNGVVKDVF-----AFGGAVENPEYLTPOGGAAPQ 1212
```

```
Db 1011 DLQAVDNPYHN-----APNGQPKAEDEYVNEPLYLNTFANTLENAEYL-----K 1055
Qy 1213 PHPPAPSPAFNDLYYWDQPPPERGA--PPSTFKGTPT-----AENPEYL 1255
Db 1056 NNLPKAKKAFNDPDYWNHSLPFRSTLQHPDLYLOEYSTKYFKQNGRIRPIVAENPEYL 1114

RESULT 5
Q9YH40
ID O9YH40; PRELIMINARY; PRT; 1165 AA.
AC O9YH40;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Receptor tyrosine kinase proto-oncogene.
GN XMRK.
OS Xiphophorus xiphidium.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX NCBI_TaxID=8086;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIO PURIFICATION;
RX MEDLINE=98241172; PubMed=9582016;
RA Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,
RA Altschmid J., Scharl M.;
RT "Activation of the Xmrk proto-oncogene of Xiphophorus by
RT overexpression and mutational alterations.";
RL Oncogene 16:1681-1690(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RIO PURIFICATION;
RA Scharl M.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U53471; AAD10500.2; -.
DR HSP; P11362; 1FGK.
DR InterPro; IPR000345; CytC heme bind.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 2.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1165 AA; 129614 MW; 7F7E38D8771A74E CRC64;

Query Match 39.2%; Score 2683; DB 13; Length 1165;
Best Local Similarity 44.9%; Pred. No. 2.7e-194;
Matches 577; Conservative 164; Mismatches 385; Indels 158; Gaps 32;

Qy 1 MELAALCRWGLLALLPPG-AAST-----QVCTGDMKRLRPASPETHLDMRLHYQCGOV 55
Db 4 LELLEL-----LALLLSIGRCCSTDPDRKVCQGTSNQMTM---LDNHLYLKMKNYSGCNV 56

Qy 56 VQGNLELTPLPTNASLFDIOIEV-----QFNFTVSWFLRPVKYSASHLEQRLRIVR 109
Db 57 VLENLEITYTQHMODLSFLOSIEQGVGYLIAMNEVST-----IPLVN-----LRLIR 104

Qy 110 GTQLFEDNYALAVLDNGDPLNNTTPTVTGASFGGLRELQLRSLTEILKGGVLIQRPOLCY 169
```

```
Db 105 GONLYEGNFTLLVMSNYOK- NPSSP- -DVOVGLKQLQLSNLTSLSGVKVSHNPLCN 161
Qy 170 QDTILWKDI FHNQALATLIDTRSRACHPCSPMKGSRGWGSESDCQSLTRTVGAGG 229
Db 162 VETINWMDIVDKTNSPTNPHAFERQCRQKDCPCVNGSCWAFPGPQHCQKFTKLLCAEQ 221
Qy 230 C-ARCKGPLPTDCHEQCAAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTRESM 288
Db 222 CNRRCRGPKPIDCCNEHCAGCTGPRATDCLACRDFDNDGCKTCTCPPPKYIDIVSHQVV 281
Qy 289 PNPEGRYTFGASCVTACPYNLSTDVGSCITLVCPHLNQEVTAEDTQRCERCKSPCARVC 348
Db 282 DNPNIKYFGAACVKECFSNVVTIE-GACVRSAGMLEVD-ENGKRSCKPCDGVCPKVC 339
Qy 349 YGLGMEHREVRVAVTSANIQFAGCKKIFGSLAFIPESFDGDPASNTAPLOPEQLVPET 408
Db 340 DGIIGTSLNTIAVNSTNIGSFNSCTKNGIILNRNSFEGDPHYKIGFMDPEHLNLT 399
Qy 409 LEEITGYLYISAWPDSLPLDSVFONLQVIRGILHNGAYS-LTQGLGISMGLRSREL 467
Db 400 VKEITGYLYVWVWENMTSLSFVQNLLEIRGRTTFRSGFSFVVQVSHLQWGLSSLKEV 459
Qy 468 GSGLALIHNTLCFVHTVPDQLFRNPHQALHTANRPEDECYGEGLACHQLCARGHCW 527
Db 460 SAGNVILKNTPLQRYASTINRRFRSDEQSIYDART-----ENQTCNECEDGCW 512
Qy 528 GPGPTQVCNCSQFLRGQCEVCEKRVLQGLPREYVNRHCLPCHPECPQNGSVTCFGPEA 587
Db 513 GPGPTMVCSLHVRGGRGVASCNLLQGEPEAOVDGRVCQCHOECLVQTDLSLTCYGGP 572
Qy 588 DQCVACAHYKPPFCFVARCPGKVPDLSPYMPKWFPPDEGACOPCINCTHSCVDLDDKG 647
Db 573 ANCSKCAHFQGPQICPRCPHMLGDGTL-IWKYADKMGQCPQCHQCTGCGSGPLSG 631
Qy 648 CPAQRASPLTSIVSAVYGLLVVGLVGVFGLIKRQCKIRKYMRLLELQETLVEPLT 707
Db 632 CRGD-IVSHSLAVGLVSGLLITVALLIVLLRRRIK-RKRTIRLLQEKELVEPLT 689
Qy 708 PSGAMPNOAQMRILKETELRKVKVLGSGAFGTVYKGIWPDGENKIPVAIKVRENTSP 767
Db 690 PSGAPNOAFLRIKETEFKDRVLGSGAFGTVYKGLWNPDPENIRIPVAIKVRENTSP 749
Qy 768 KANKEILDEAYMAGVGPVYRLLGICLTSTVOLATOLMPYGCILDHVRNRLGSGD 827
Db 750 KNGEVLDEAYMASVDHPHVCRLGICLTSAVLVQLTQMPYGCILDHVRQERICQW 809
Qy 828 LLNMCQIAKMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADG 887
Db 810 LLNMCQIAKGMNYLEERHLVHRDLAARNVLKPNHVKITDFGLSKLLTADEKEYQAHG 869
Qy 888 GKVPKMWALESLRRRTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGBRL 947
Db 870 GKVPKMWALESLIOWTTHQSDVWSYGVTVWELMTFGSKPYDGIIPAKIEIASVLNGBRL 929
Qy 948 POPPTCTIDVMYKMWIDSECPREELVSEFSRMARDPQRFVVIQNEDLGPASPLD 1007
Db 930 POPPTCTIEVMIILKMWIDPSRPRELVEFSQARDPSRYLVVQG---NLPSPSD 986
Qy 1008 STFVRSLLDEDDMDGLVDAAEYLVPOQGFPCDPAPGAGMVHHRHRSSTRSGGDLTL 1067
Db 987 RRLFSRLSSDD--DVVDADAYLL-----RYKRIN-RQGS----- 1018
Qy 1068 GLEPSEEARPSPLAPSGAGSDVPDGLGMAKGLQSLPHTDPSLQRYSEDPV-PL 1126
Db 1019 -----EPCIPPNGH-----FVRENSIALRYSIDPTQNAL 1047
Qy 1127 PSETDGVAPLTCSPQPVVNPQDVRPQP-----PSPRE-----GLP-AARPAGATL 1173
Db 1048 EKDLGDH-----EYVQPGSETSRSLSDIYNPNYEDLTDGQGVSLSSQAEATNF 1097
Qy 1174 ERAKTLSPGKNGVKNVDVFAFGAVENPEYLTPOGGAAPQPPPPAFSPAFNLVYWDQDP 1233
```

```
Db 1098 SRPEYLTNTQNSL---PLVSSGSMDDPDY---QAG-----YQAAF-----L 1132
Qy 1234 PERGAPPSTFKGTPTAENPEVIGL 1257
Db 1133 POTGALTGNMGFUPAAENLEYLGL 1156
RESULT 6
P79754
ID P79754 PRELIMINARY; PRT; 1328 AA.
AC P79754;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ErBB3.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99177347; PubMed=10077531;
RA Gellner K., Brenner S.;
RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
rubripes.";
RL Genome Res. 9:251-258(1999).
DR EMBL; AF056116; AAC34391.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR ATP-binding; Transferase.
SQ SEQUENCE 1328 AA; 148613 MW; A333039258B647E9 CRC64;
```

```
Query Match 33.1%; Score 2263; DB 13; Length 1328;
Best Local Similarity 39.9%; Pred. No. 2.3e-162;
Matches 517; Conservative 157; Mismatches 418; Indels 204; Gaps 34;
Qy 9 WGLLLALLPP--GAASTQ---VCTGTDMLRLPASPTHLDMLRLHYQGCVVQGNLEL 62
Db 4 WRLTLMCVASRLRAASQTQEA VCPGQNGLSSTGSGENQVNLNKDKYKGCETMGNLEI 63
Qy 63 TYLPTNASLFLQDIOEQVQNNFTVSEFLRVKVSASHLEQ-----RLRIVRGTLQFEDNY 118
Db 64 TQIESNWDPSFLKTIREV-----TGVVL---IAMNHQEIPLGLRVIRGNSLYERRF 113
Qy 119 ALAVLDNGDPLNNTPTVGTASPGGLRELQLRSLTEILKGGVLIORNPOLCYQDTILWKDI 178
Db 114 ALSVFLN-----YPKDG--PSGLNQLGLMNLTEILDGGVQIINNKYLRYPVWYWRDI 164
Qy 179 FHKNQLALTLIDNRSRACHPCSPMKGSRGWGSESDCQSLTRTVGAGG-ARCKGPL 237
Db 165 I-RNNDAPIEIQFNGRGVCH---KSC-GNYCWPFGDKQQLITKTVCAPQCNDRCFGTS 219
Qy 238 PTDCCHQCAAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESNPPEGRYTF 297
Db 220 PRDCHIECAAGCKGPLDTCFACRLFNDGACVPQCPQTLIYNKQTFQMETNPAKYOY 279
Qy 298 GASCVTACPNYLSSTDVGSCTLVCPHLNQEVTAEDGTQR-CEKSKPCARVCYGLGMEHL 356
Db 280 GSICVSCQPTHFV-VDGSSCVSVCPDPKMEV--ERGSQRCQELCSGLCPKVCETGAE-- 334
```



```
Qy 489 DQFRNPHQALLHTANRPEDECVEGLACHQLCARGHGWGPGTQCVCNCSQFLRGQECVE 548
Db 444 SEIKKSDHVMVQKRNATECHBEGMECSQCSKAGCWGKGPQCLECKNVKYGKCLD 503
Qy 549 ECRVLQGLPRY-VNARHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPCVARCP 607
Db 504 SKC---SLPRLYSVDSKTCGCHQCKD-----FCYGNEDNCGSCMNVDKGRFCVAEC 555
Qy 608 -----SGVKPDLVYMPIKFPD-- 624
Db 556 TTKHAMNGTCINCHKTCVCGRPRDTIADGCIISCDKAIIGSDAKIERCLMKDSCPDGY 615
Qy 625 -----EEG----- 627
Db 616 YSDVVLQEEGLKQLSGKAVCRKCHPRCKKCTGYGFHEQFCOEQCYGKKGECDECPQD 675
Qy 628 -----ACQPCINCT-----HSCVDL-----DD-----KGCRAEO--- 652
Db 676 FYANETRILCLPCHQECRGCHGLGDHHECRNLKLFEGDPYDNNATTFTCVSNCPASHPYK 735
Qy 653 -----RASPLTSIVSAVVGILLVVVLGVFGI---LKKRQOK 687
Db 736 RFPQAGKIGYCSADSNQSGRLIEPQTKIVMGVSNALILLVCVFGIAFVLSRHNK 795
Qy 688 IRKYMTRRLLOETELVEPLTPSGAMPNOAQRILKETELRKVKVLGSGAFVYKGIWIP 747
Db 796 KDAVMTMALAGCEDSEPLRSNVGNLTKRIIEAIEIRGGVGLMGAFGRVFGVWMP 855
Qy 748 DGENVKIPVAIKVIRENTSPKANKEILDEAVYMGVSPYVSRLLGICLTSTVOLVQLM 807
Db 856 EGESVKIPVAIKVLMESGSSSEKFELEAYIMASVEHPNLLKLLAVCMTSQMMILITQLM 915
Qy 808 PYGLLDHVRNCRGLSODILNNCMOIAKMSVLEOVLRVHRLDAAARNVLVKSFNHVKI 867
Db 916 PLGCLLDYVRNKKDKIKSALNNWSTQIAREGMAYLEBERRLVRHRLDAAARNVLQVTPSCVKI 975
Qy 868 TDFGLARILDDIETEIADGGKVPKIMWALSESILRRFTTHOSDVMSYGVTVWELMTFGAK 927
Db 976 TVFGLAKLLDFSDRYAAGGKMPKMLALECIRHRVFTSKSDVWAFGIITWELLTYGAR 1035
Qy 928 PYDGIAREIPDLLEKGERLPPOPICTIDVIMVWKMDISECRPRELVSFSRMAR 987
Db 1036 PYENVPAKDVPELIEIGHKLPQDICSLDVYVYLLSCWVLDADARPTFKQLAETFAEKAR 1095
Qy 988 DPQRFVVIQNEIDLGPASPLDSTFYRSLLDDMDGLV----- 1024
Db 1096 DPGRYLMI-----PGDKFMLPLPSYTNODEKDLIRTLAPVAMAAAAAAGASNV 1146
Qy 1025 -----DAEEYLVPOGFFCPDPAPGAGGMVHRRHSSTRSGGDLTLGLEPSEEBAPR 1078
Db 1147 VPSTIAETDEYLPKTRPSIMLPGSA-----VEPS-DEMPK 1182
Qy 1079 S-----PLAP-----SEGAGSDVFGDLGMGAAGKGLQSLPHTDPSLORYSEDPVPLPSE 1129
Db 1183 SLRYCKDPLKPDDETGDGKEV-----GVGGIR-----LNLPLD 1216
Qy 1130 TDGVVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAGATLERAKTLSPKNGVVKD 1189
Db 1217 EDDYLP-TCOSQ-----NQS-----TPG-----YMD 1237
Qy 1190 VFAFGGAVENPEYL-----TFQGAAPQPHPPAFSPAFDNLVYWDQDPPRGAPP 1240
Db 1238 LIGVPASVDNPEYLMGSGTQAAGLAGQSMG--PHTPP-----PP 1274
Qy 1241 STFGTPTAENPE 1253
Db 1275 NTPNGMETHQHSQ 1287
RESULT 8
Q9UK79
ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
```

```
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=99415951; PubMed=10485918;
RX Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RA "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).
[2]
RN SEQUENCE FROM N.A.
RP Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF177761; AAD56009.2; -
DR InterPro; IPR004049; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FEC1BE347E2D030C CRC64;

Query Match 25.9%; Score 1775.5; DB 4; Length 419;
Best Local Similarity 85.0%; Pred. No. 3.8e-126;
Matches 340; Conservative 9; Mismatches 38; Indels 13; Gaps 4;

Qy 1 MELAAALCRWGLLALLPPGAASQVCTGTDMLKRLPASPTHLDMLRHLYGCGVVOGNNL 60
Db 1 MELAAALCRWGLLALLPPGAASQVCTGTDMLKRLPASPTHLDMLRHLYGCGVVOGNNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQ--GYVLIHNRVQVPL----QRLRVRGTLQDFEDNYAL 120
Db 61 ELTYLPTNASLSFLQDIQEVQ--GYVLIHNRVQVPL----QRLRVRGTLQDFEDNYAL 114
Qy 121 AVLNDGDPNNTPTVTGASPGGLREQLRSLTILKGGVLIQRNPQLCYQDTILWKDIFH 180
Db 115 AVLNDGDPNNTPTVTGASPGGLREQLRSLTILKGGVLIQRNPQLCYQDTILWKDIFH 174
Qy 181 KNNQLALTLDITNRSRACHPCSPMCKSGRCWGSSEDCQSLTRTVACAGGCAKCKGPLPTD 240
Db 175 KNNQLALTLDITNRSRACHPCSPMCKSGRCWGSSEDCQSLTRTVACAGGCAKCKGPLPTD 234
Qy 241 CCEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 300
Db 235 CCEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 294
Qy 301 CVTACPNYILSTVGSGCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 360
Db 295 CVTACPNYILSTVGSGCTLVCPLNHNEVTAEDGTQRCCKSKPCAR----GTHSLPRPP 349
Qy 361 AVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQ 400
Db 350 AAVPVPLRMQPG--PAHPVLVSFLRPSMDLVSAFYSLPLAP 387

RESULT 9
Q8R2X1
ID Q8R2X1 PRELIMINARY; PRT; 367 AA.
AC Q8R2X1;
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
```

```

RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RW EMBL; BC027080; AAH27080.1; -
KW Hypothetical protein.
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match 25.4%; Score 1739; DB 11; Length 367;
Best Local Similarity 88.0%; Pred. No. 1.8e-123;
Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

Qy 895 MALESILRRRTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGRLLPQPPICT 954
Db 1 MALESILRRRTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGRLLPQPPICT 60

Qy 955 IDVTWIMVKWMIIDSECRPRELSEFSESRWARDPQREVVIONEDLGPASPLDSTFYRSL 1014
Db 61 IDVTWIMVKWMIIDSECRPRELSEFSESRWARDPQREVVIONEDLGPSSPMDSSTFYRSL 120

Qy 1015 LEDDDMGDLVDAEYLVPOQGFPCDPAPGAGGMVHRRSSSTRSGGGDLTLGLPSEE 1074
Db 121 LEDDDMGDLVDAEYLVPOQGFPCDPALGNGSTAHRRHRSARSAGGELTLGLPSEE 180

Qy 1075 EAPRSPLAPSEAGSDVFDGLGMAKGLQSLPHTDPSPLQRYSEDPVPLPSETDGYV 1134
Db 181 EAPRSPLAPSEAGSDVFDGLGMAKGLQSLPHTDPSPLQRYSEDPVPLPSETDGYV 240

Qy 1135 APLTCSPOPEVNOPDVRPQPSREGPLPAARAGATLERAKTSLSPKNGVWVDFAFG 1194
Db 241 APLTCSPOPEVNOPDVRPQPSREGPLPAARAGATLERAKTSLSPKNGVWVDFAFG 300

Qy 1195 GAVENPEYLTPOGGAAPHPHPPAFSPAFDNLYYWDQPPRRGAPPSTFKGTPTAENPEY 1254
Db 301 GAVENPEYLA PRAGTASQPHSPAFSPAFDNLYYWDQNSSESGPPSTFKGTPTAENPEY 360

Qy 1255 LGLDVPV 1261
Db 361 LGLDVPV 367

RESULT 10
Q86712 ID Q86712 PRELIMINARY; PRT; 729 AA.
AC Q86712;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Polyprotein.
GN POLYPROTEIN.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60725.1; -
DR HSP; P03322; I46S.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004028; Retro_M.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02813; Retro_M; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.

```

```

DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EFE1D63 CRC64;

Query Match 25.1%; Score 1720; DB 15; Length 729;
Best Local Similarity 54.8%; Pred. No. 1.4e-121;
Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15;

Qy 575 PQNGSVTCFGEADQCVACAHYKDPFPCVACRPSGVKPDLSYMPIMKFPDDEGACQPCPI 634
Db 141 PEETATPKTGP--DHCNKKCAHFIDGPHCVKACAGVLGENDTL-VWKYADANAVCOLCHP 197

Qy 635 NTHSCVDLLDKGCPAQBRASPLTISVAVV-GILLVVLGVVFGILIKRQOKIRKYTM 693
Db 198 NCTRGCKGPGLEGCP--NGSKTPSIAAGVVGGLLCLVWVGLGILGLYLRER-HIVKRTL 253

Qy 694 RRLQETELVEPLTPSGAMNQOMRILKETELRKVKVLGSGAGTGVYKGIWIPDGENVK 753
Db 254 RRLQERELVEPLTPSGEAPNOAHLRLKETEFKVKVLGSGAGTGVYKGLWIPEGEVK 313

Qy 754 IPVAIKVLRRENTSPKANKEILDEAYVMAGVSPVSRLLGLCLTSTVOLVTLQMPYGCLL 813
Db 314 IPVAIKELRENTSPKANKEILDEAYVMASVDNPRVCELLGLCLTSTVOLVTLQMPYGCLL 373

Qy 814 DHVRENRLG3QDOLLNWCQIAKGM5YLEDVRLVHRDLAARNVLKSPNHVKITDFGLA 873
Db 374 DYIREHKDNTG3QVLLNWCQIAKGM5YLEDVRLVHRDLAARNVLKTPQHVKITDFGLA 433

Qy 874 RLDDIDETEHAGGKVPKIMMALESILRRRPTHQSDVWSYGVTVWELMTFGAKPYDGIP 933
Db 434 KLLGADEKEVHAGGKVPKIMMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIP 493

Qy 934 ARIPDLLEKGERLPQPICTIDVTMTVMKWMIDSCRRPRFELVSEFSRMARDPORFV 993
Db 494 ASHSSVLEKGERLPQPICTIDVTMTVMKWMIDSCRRPRFELVSEFSRMARDPORFV 553

Qy 994 VIQ-NEDLGPASPLDSTFYRSLLEDGDLVDAEYLVPOQGFPCDPAPGAGGMVHHR 1052
Db 554 VIQGERMHLPSPTDSKFYRTLMEEDMEDIVDAEYLVPHQGF----- 598

Qy 1053 HRSSTRSGGGDLTLGLEPSEEA PRSPL-----APSEAGSDVFDGDLGMAKGLQSL 1107
Db 599 -NSPST-----SRTPLLSLSATSNSATNCID-----RNGQGH 631

Qy 1108 PTHDPSPLQRYSEDPVPLPSET--DGYVAPLTCSPQPEVNOPDVRPQPSREGPLPA 1165
Db 632 PVREDSFVQRYSSDPTGNFLEESIDGFL-----PAPEYVQ--LMPKKPS----- 675

Qy 1166 ARPAGATLERAKTSLSPKNGVWVDVF-----AFGGAVENPEYL 1203
Db 676 -----TAMVQNOIYNNISLTAISKLPMDSRVQNSHSTAVDNPEYL 715

RESULT 11
Q86714 ID Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";

```



```

Qy 943 KGERLPPPICTIDYIMVIMVCKWIDSECRPRFRELSEYSEFMRMARDPQFVVIQ-NEDLG 1001
Db 354 KGERLPPPICTIDYIMVIMVCKWIDSECRPRFRELSEYSEFMRMARDPQFVVIQ-DERMH 413
Qy 1002 PASPLDSTFYKSLLEDMDGLVDABEYLVPOQGFCCPDAPAGAGMVMHRRHSSTRSG 1061
Db 414 LPSPTDSKFYRTLMEEDMEDIVDAEYLVPHQGF--NSPST--- 454
Qy 1062 GGDITLGLPESEAPRSPPL-----APSEGAGSDVFDGDLGMAKGLQSLTHDPPLQ 1116
Db 455 -----SRTPLLSSLSATSNNSATNCIDRNG--H----- 481
Qy 1117 RYSDPTVPLPSETGTYVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPACAT-LER 1175
Db 482 -----PVREDFGL-----PAPEYVNQ--LMPKKPTAMVQIQIYVLSLTAISK 523
Qy 1176 AKTISPCKNGVYKOVFAFGGAVENPEYL 1203
Db 524 LPMDSRVQN-----SHSTAVDNPPEYL 544

RESULT 15
Q9WVF5
ID Q9WVF5 PRELIMINARY; PRT; 655 AA.
AC Q9WVF5;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor
DE isoform 3).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C.,
RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Maible N.J.;
RT "Alternative Transcripts from the Human and Mouse EGFR Genes Encode
RT Carboxy-Terminal Truncated Receptors.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTA;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sinclair C.S., Pearshall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikola I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

```

Search completed: July 22, 2003, 09:00:11
Job time : 53.5887 secs

```

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AF124513; AAD4149.1; -.
DR EMBL; AF275366; AAG28047.1; -.
DR EMBL; AF275364; AAG28047.1; JOINED.
DR EMBL; AF275365; AAG28047.1; JOINED.
DR EMBL; AK004944; BAE23688.1; -.
DR EMBL; AK004883; BAE23641.1; -.
DR EMBL; AK004911; BAE23662.1; -.
DR MGD; MGI:95294; Egr.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR SMART; SM00261; FU; 3.
KW Receptor.
SQ SEQUENCE. 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

```

```

Query Match 21.4%; Score 1463.5; DB 11; Length 655;
Best Local Similarity 42.7%; Pred. No. 3.3e-102;
Matches 276; Conservative 101; Mismatches 238; Indels 31; Gaps 9;

Qy 11 LLLALLPPGAA--STQVCTGTDMLRLPASPETHDMLRLHLYOGCQVVOGNELEYLPTN 68
Db 14 LLLTALCAAGGALBEKVCQGTSNRLTQGTGFEDHFLSLQRMNVCVVLGNLEYTYVORN 73
Qy 69 ASLSFLQDIOEVO-FNNFTVSFWLRVPKVSASHLEQRRLRVRGTLQFPEDNVAVLNDGD 127
Db 74 YDLSFLTKTIOEVAGYVLIANTVERIPL-----ENLIIRGNALYENTYALAILSN-- 124
Qy 128 PLANTTPVTGASPGGLRELQRLSLTEILKGVLIQIRNPQLCYODTILWKDI----FKNN 183
Db 125 -----YGTNRTGLRELPMRLQEILIGAVRFSNNPILCNMDTIQWRDIVQVFNMSNM 176
Qy 184 QLAITLIDTRSRACHPCSPKSGRCWGESSEDCQSLTFTVCAGGCA-RCKGPLPTDCC 242
Db 177 SMDL----QSHPSSCPDKPCSPNGSCWGGEGENCQKLTIIQAQCSHRCGRSPSDCC 232
Qy 243 HECCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMENPEGRYTFGASCV 302
Db 233 HNQAAGCTGPRESDCLVCOCKFODEATCKTCTPLMLYNTTYQMDVNPBGKVSFGATCV 292
Qy 303 TACFYNLYSDVGSCTLVCPHNOEVTAEDGTORCEKSKPCARVCYGLGMEHLREVAV 362
Db 293 KKCPRYNVVTDHSGVRCACGPDYVEV-EEDGIRKCKKCDGCRKVCNGIGIGEFKDTLSI 351
Qy 363 TSANIQEPACKKIFGSLAPLPESFDGDPASNTAPLOPEQLQVPELEETGLYLYISAMP 422
Db 352 NATNIKHFKYCTAISGLHILPVAFGDSFTRTPPLDPRELEILKTVEITGFLIIQAMP 411
Qy 423 DSLPDLISVFONLOVIRCIHNGAYSLLTLOGLGISWGLRSLRSLRELGSGLAIHHNTHLCF 482
Db 412 DNTDLHAFENLEIRGTRKHQGFSLAVVGLNITSLRSLRSLKEISDGVIIISGNRLCY 471
Qy 483 VHTVPMDQLFRNPQALLHTANPEDECVGEGELACHOLCARGHCWGPGPTOCVNCOSFLR 542
Db 472 ANTINWKKLFGTENQTKIMNNAEKDKAVNHVNCNPLSCSSEGCWGPDEPRDCVSNVR 531
Qy 543 GQECVECRVLOGLPREYVNAHCLPCHPECOQNGSVTCFGEADOCVACAHYKDPFPC 602
Db 532 GRECVKCNLTLEGEPREFVNSCICQHPCLPQAMNITCTGRGPDNCIOCAHYIDGPHC 591
Qy 603 VARCPGSKVDLSYMPYWKPFDEEGACQPCPINCTHSCVDLDKGC 648
Db 592 VKTCPAGIMGNVTL-VWKYADANNVCHLCHANCTYCCAGPGGLQGC 636

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 38.1589 Seconds
(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-59-73-14

Perfect score: 6947

Sequence: 1 MELAALCRWGLLLALLPPG.....TFKGTPTAENPEYGLDVPV 1261

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6713	98.0	1255	21	Human heregulin 2
2	6713	98.0	1255	22	Human heregulin 2
3	6713	98.0	1255	22	HER2 transgene pla
4	6713	98.0	1255	23	Human HER2 (ErbB2)
5	6707	98.0	1255	17	HER-2/neu protein.
6	6707	98.0	1255	20	Human HER-2/neu on
7	6707	98.0	1255	21	Human HER-2/neu pr
8	6707	98.0	1255	21	Amino acid sequenc
9	6707	98.0	1255	22	Human HER-2/neu pr
10	6707	98.0	1255	22	HER2/neu amino aci

11	6707	98.0	1255	23	AAE24067	Human Her-2 protei
12	6707	98.0	1255	23	AAE20479	Human Her-2/neu pr
13	6707	98.0	1255	23	AAE51143	Human Her-2/neu pr
14	6707	98.0	1255	23	AAU77114	Human Her-2/neu po
15	6664	97.3	1433	14	AAE39568	Sequence of c-erbB
16	6543	95.6	1223	23	AAU98923	Human breast cance
17	6390	93.3	1200	21	AAE21208	Rat HER-2/neu pr
18	5911.5	86.3	1256	21	AAE21199	Rat HER-2/neu pr
19	5911.5	86.3	1256	23	AAE51144	Rat Her-2/neu onco
20	5891.5	86.0	1256	21	AAE21206	Mouse Her-2/neu pr
21	5891.5	86.0	1256	22	AAE62860	Amino acid sequenc
22	5891.5	86.0	1256	23	AAE51151	Mouse Her-2/neu on
23	4793	70.0	919	23	AAE21203	Human HER-2/neu fu
24	4793	70.0	919	23	AAE51148	Her-2/neu extracel
25	4041.5	59.0	920	23	AAE51152	Mouse Her-2/neu ex
26	4041.5	59.0	926	23	AAE51153	Mouse Her-2/neu ex
27	3677	53.7	712	21	AAE21204	Human HER-2/neu fu
28	3677	53.7	712	23	AAE51149	Her-2/neu extracel
29	3531	51.6	782	18	AAW19764	Her2-GM-CSF immuno
30	3529	51.5	653	21	AAE21200	Extracellular HER-
31	3529	51.5	653	23	AAE51145	Human Her-2/neu on
32	3491	51.0	645	22	AAE60408	Human ErbB2 oncopr
33	3491	51.0	645	22	AAE61593	Human ErbB2 extrac
34	3426	50.0	951	21	AAE44993	DC8scFV-erbB2EC fu
35	3323	48.5	624	11	AAE08222	Extracellular port
36	3098	45.2	1210	21	AAE19259	Amino acid sequenc
37	3098	45.2	1210	21	AAE50616	Human EGF receptor
38	3098	45.2	1210	23	AAE23019	Human Her-1 protei
39	3098	45.2	1210	23	AAE50768	Human epidermal gr
40	3096	45.2	1210	22	AAE68420	Amino acid sequenc
41	3084	45.0	583	23	AAE20483	Human protein for
42	3084	45.0	587	23	AAE20481	Human protein for
43	3083	45.0	589	23	AAE20484	Human protein for
44	3083	45.0	600	23	AAE20482	Human protein for
45	3057	44.6	1210	23	ABF51768	Human epidermal gr

ALIGNMENTS

RESULT 1
AAE2620
ID AAY92620 standard; Protein; 1255 AA.
XX
AC AAY92620;
XX
DT 10-AUG-2000 (first entry)
XX
DE Human heregulin 2 (Her2).
XX
KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
XX
OS Homo sapiens.
XX
PH Key
FT Domain
FT Location/Qualifiers
FT /label= N-terminal
FT /note= "mature polypeptide"
FT 5...25
FT Region
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 59...73
FT Region
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 103...117
FT Region
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 149...163
FT Region
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 174...323
FT Domain

FT Region /label= Cysteine_rich_domain
 FT 210..224
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 250..264
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 324..483
 FT /label= Ligand_binding_domain
 FT 325..339
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 369..383
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 465..479
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 484..623
 FT /label= Cysteine_rich_domain
 FT 579..593
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 624..654
 FT /label= Transmembrane_domain
 FT 632..652
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 653..667
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 655..1010
 FT /label= Tyrosine_kinase_domain
 FT 661..675
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 695..709
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 710..730
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 1011..1235
 FT /label= C-terminal_domain
 XX WO200020027-A2.
 XX 13-APR-2000.
 XX 05-OCT-1999; 99WO-DK00525.
 XX 05-OCT-1998; 98DK-0001261.
 XX 20-OCT-1998; 98US-0105011.
 XX (MEBI-) M & E BIOTECH AS.
 XX Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
 PI Gautam A, Birk P, Karlsson G;
 XX WPI; 2000-349917/30.
 DR N-PSDB; AAA09455.
 XX Inducing immune responses to weakly immunogenic, tumor associated
 PT peptide antigens for the treatment of breast and prostate cancer
 XX Claim 62; Page 193-198; 220pp; English.
 XX This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
 CC Her2 can be used in the claimed method as an autovaccine to induce a CTL
 CC response. Subdominant CTL epitopes, antibody binding regions and
 CC cysteine residues involved in disulfide bonds are preserved in the
 CC immunogenized forms. Regions suitable for the insertion of foreign T
 CC helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
 CC cell-associated peptide antigens (PA) such as those associated with
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
 CC The method comprises effecting simultaneous presentation by antigen
 CC producing cells (APCs) of the animals immune system of: (1) at least 1
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
 CC B-cell group derived from the cell-associated PA; and (2) at least 1
 CC first T helper cell group which is foreign to the animal. Analogues of
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
 CC part of all known and predicted CTL and B-cell epitopes of the respective
 CC PA and including at least one foreign T helper epitope are also claimed.
 CC The method is used to treat prostate, prostate/breast or breast cancer
 CC when the PA is human PSM, FGF8b and Her2, respectively.

XX Sequence 1255 AA;

Query Match	98.0%;	Score 6713;	DB 21;	Length 1255;
Best Local Similarity	98.5%;	Pred. No. 0;		
Matches 1242;	Conservative	4;	Mismatches	9;
			Indels	6;
			Gaps	2;

Qy 1 MELAALCRWGLLLALLPPGAASQTQCTGDMKRLPASPETHDMLRHLVGGCQVQGNL 60
 Db 1 MELAALCRWGLLLALLPPGAASQTQCTGDMKRLPASPETHDMLRHLVGGCQVQGNL 60
 Qy 61 ELTYLPTNASLSFLQDIQEVQFNFTVSFWLRVPKVSASHLEQRIRIVRGQTLPEDNYAL 120
 Db 61 ELTYLPTNASLSFLQDIQEVQ--GYVLIANQVRQVPL----QRLRIVRGQTLPEDNYAL 114
 Qy 121 AVLNDGDLNNTTPTVTGASPGGLRELQRLSLTELKGGVLIQRNPQLCYQDTILWKDIFH 180
 Db 115 AVLNDGDLNNTTPTVTGASPGGLRELQRLSLTELKGGVLIQRNPQLCYQDTILWKDIFH 174
 Qy 181 KNNQALTLIDTNRSRACHPCSPMCKGSRGWGSESDCQSLTRTVGAGGCARCKGPLPTD 240
 Db 175 KNNQALTLIDTNRSRACHPCSPMCKGSRGWGSESDCQSLTRTVGAGGCARCKGPLPTD 234
 Qy 241 CCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPPEGYRTFGAS 300
 Db 235 CCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPPEGYRTFGAS 294
 Qy 301 CVTACPNYLSLTDVGSCTLVCPLNHQNVEAEDGTQRCCKSKPCARVCYGLGMEHLREVR 360
 Db 295 CVTACPNYLSLTDVGSCTLVCPLNHQNVEAEDGTQRCCKSKPCARVCYGLGMEHLREVR 354
 Qy 361 AVTSANTQEFAGCKKIFGSLAFIPESPDGPASNTAPLOPELOQVFTBELBITGLYISA 420
 Db 355 AVTSANTQEFAGCKKIFGSLAFIPESPDGPASNTAPLOPELOQVFTBELBITGLYISA 414
 Qy 421 WPDSLPDLVSFQNLQVIRGRILHNGAYSLLTQIGISWLGRLSRLRELSGLALIHNNTHL 480
 Db 415 WPDSLPDLVSFQNLQVIRGRILHNGAYSLLTQIGISWLGRLSRLRELSGLALIHNNTHL 474
 Qy 481 CFVHTVPWDQLFRNPQHALLHTANRPEDECVGEGEGLACHQLCARGHCGPGTQCVCNSQF 540
 Db 475 CFVHTVPWDQLFRNPQHALLHTANRPEDECVGEGEGLACHQLCARGHCGPGTQCVCNSQF 534
 Qy 541 LRQGEVCEERVLQGLPREVYNARHCLPCHPECPQNGSVTCFGEADQCACAHYKDP 600
 Db 535 LRQGEVCEERVLQGLPREVYNARHCLPCHPECPQNGSVTCFGEADQCACAHYKDP 594
 Qy 601 FCVARCPSGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKCPAQRASPLTSI 660
 Db 595 FCVARCPSGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKCPAQRASPLTSI 654
 Qy 661 VSAVVGILLVVVLGWVFGILIKRRQKIRKYTMERLLQETELVEPLTPSGAMPNQAQMRI 720
 Db 655 VSAVVGILLVVVLGWVFGILIKRRQKIRKYTMERLLQETELVEPLTPSGAMPNQAQMRI 714
 Qy 721 LKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKULRENTSPKANKEILDEAYVM 780
 Db 715 LKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKULRENTSPKANKEILDEAYVM 774

Qy 781 AGVSPVSRLLGICLTSTVOLTPQLMPYGCCLLDHVRNRRLGSDQLLNWCMIKAGMS 840
 Db 775 AGVSPVSRLLGICLTSTVOLTPQLMPYGCCLLDHVRNRRLGSDQLLNWCMIKAGMS 834
 Qy 841 YLEDRVLRHDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWKMALES 900
 Db 835 YLEDRVLRHDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWKMALES 894
 Qy 901 LRRRFTHQSDVWSYGVTVWELMTGAKPYDGPAREIPDLLEKGERLPQPPICITIDVYMI 960
 Db 895 LRRRFTHQSDVWSYGVTVWELMTGAKPYDGPAREIPDLLEKGERLPQPPICITIDVYMI 954
 Qy 961 MVKCMWIDSECRPRFRELVSFMSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1020
 Db 955 MVKCMWIDSECRPRFRELVSFMSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1014
 Qy 1021 GDLVDAEYLVPOQGFCCPDPAAGAGGVHHRHSSSTRSGGDLTLGLEPSEEAAPRSP 1080
 Db 1015 GDLVDAEYLVPOQGFCCPDPAAGAGGVHHRHSSSTRSGGDLTLGLEPSEEAAPRSP 1074
 Qy 1081 LAPSEGAGSDVDFDGLGMAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCS 1140
 Db 1075 LAPSEGAGSDVDFDGLGMAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCS 1134
 Qy 1141 POPEYVNPQDVRPOPSPREGPLPAAPAGATLERAKTTLSPGKNGVVKDVPFAGGAVENP 1200
 Db 1135 POPEYVNPQDVRPOPSPREGPLPAAPAGATLERAKTTLSPGKNGVVKDVPFAGGAVENP 1194
 Qy 1201 EYLTPOGGAAPQHPHPPAFSPAFNLVYWDODPBERGAPSTFKGTPTAENPEYGLDVP 1260
 Db 1195 EYLTPOGGAAPQHPHPPAFSPAFNLVYWDODPBERGAPSTFKGTPTAENPEYGLDVP 1254
 Qy 1261 V 1261
 Db 1255 V 1255

RESULT 2
 AAEL12130
 ID AAEL12130 standard; Protein; 1255 AA.
 XX
 AC AAEL12130;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Human tyrosine kinase-type receptor, HER-2.
 XX
 KW Therapeutic compound; major histocompatibility complex; vaccine; antigenic peptide; MHC; immunoregulatory; immune response; HER-2; adoptive immunotherapy; anti-cancer; breast cancer antigen; APC; antigen presenting cell; human; tyrosine kinase-type receptor.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Region 774..782
 FT /note= "Antigenic epitope"
 XX
 FT WO200168677-A2.
 XX
 FT 20-SEP-2001.
 XX
 XX 16-MAR-2001; 2001WO-US40328.
 XX
 XX 16-MAR-2000; 2000US-0527487.
 XX
 XX (GENZ) GENZYME CORP.
 PA
 XX Nicolette CA;
 PI
 XX WPI: 2001-616284/71.
 DR N-PSDB; AAD19731.
 XX

PT Novel synthetic therapeutic compound for inducing immune response and
 PT for use in adoptive immunotherapy, has enhanced binding to major
 PT histocompatibility molecules and enhanced immunoregulatory properties
 XX
 PS Claim 4; Page 63-67; 69pp; English.
 XX
 CC The invention relates to synthetic therapeutic compounds (antigenic
 CC peptides) with enhanced binding to major histocompatibility complex
 CC (MHC) molecules and enhanced immunoregulatory properties relative
 CC to their natural counterparts. Compounds of the invention are useful
 CC for inducing an immune response in a subject and for use in adoptive
 CC immunotherapy. They are useful as components of anti-cancer vaccines
 CC and to expand immune effector cells that are specific for cancers
 CC characterised by expression of the breast cancer antigen, HER-2.
 CC Polynucleotides that encode peptides of the invention are useful as
 CC hybridisation probes and as primers for the detection of genes of gene
 CC transcripts that are expressed in antigen presenting cells (APCs), to
 CC confirm transduction of polynucleotides into host cells. The present
 CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
 CC of the invention are designed based on the HER-2 antigenic peptide
 CC (774-782).

XX Sequence 1255 AA;

Query Match 98.0%; Score 6713; DB 22; Length 1255;
 Best Local Similarity 98.5%; Pred. No. 0;
 Matches 1242; Conservative 4; Mismatches 9; Indels 6; Gaps 2;

Qy 1 MEAALCRWGLLALLPPGAASQVCTGDMKRLPASPETHLDMLRHLVGGCQVQGNL 60
 Db 1 MEAALCRWGLLALLPPGAASQVCTGDMKRLPASPETHLDMLRHLVGGCQVQGNL 60
 Qy 61 ELTYLPTNASLSFLQDIQEVQFNFTVSFWLRVPKVSASHLEQRLRVRGTLQFEDNYAL 120
 Db 61 ELTYLPTNASLSFLQDIQEVQ--GYVLIANQVRQVPL---QKRLRVRGTLQFEDNYAL 114
 Qy 121 AVLDNGDPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFH 180
 Db 115 AVLDNGDPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFH 174
 Qy 181 KNNQALTLIDTNRSRACHPCSPMKGSRGWGSSSDCQSLTRTVAGGACARCKGPLPTD 240
 Db 175 KNNQALTLIDTNRSRACHPCSPMKGSRGWGSSSDCQSLTRTVAGGACARCKGPLPTD 234
 Qy 241 CCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGYTTFGAS 300
 Db 235 CCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGYTTFGAS 294
 Qy 301 CVTACPNYLSLTDVGSCTLVCPLNHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREVR 360
 Db 295 CVTACPNYLSLTDVGSCTLVCPLNHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREVR 354
 Qy 361 AVTSANIQEPAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFTLEBITGYLYISA 420
 Db 355 AVTSANIQEPAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFTLEBITGYLYISA 414
 Qy 421 WPDLSPLSVFQNLQVIRGRILHNGAYSLTLQGLIGLSWLGLRSLRELGSGLALIHNNHNL 480
 Db 415 WPDLSPLSVFQNLQVIRGRILHNGAYSLTLQGLIGLSWLGLRSLRELGSGLALIHNNHNL 474
 Qy 481 CFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGPTQCVNCSQF 540
 Db 475 CFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGPTQCVNCSQF 534
 Qy 541 LRQGECEVECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDDP 600
 Db 535 LRQGECEVECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDDP 594
 Qy 601 FCVARCPGKVPDLISYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCRAEORASPLTSI 660
 Db 595 FCVARCPGKVPDLISYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCRAEORASPLTSI 654

```
Qy 661 VSAVVGILLVVVLGVVFGILLIKRQOKIRKYVIMRLLQETELVPLTPSGAMPNQAOVRI 720
Db 655 VSAVVGILLVVVLGVVFGILLIKRQOKIRKYVIMRLLQETELVPLTPSGAMPNQAOVRI 714
Qy 721 LKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVM 780
Db 715 LKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVM 774
Qy 781 AGVSPVYSRLILGLISTVQLPMPYGLLDHVRNRRGLSQDILLNMCQIAKMS 840
Db 775 AGVSPVYSRLILGLISTVQLPMPYGLLDHVRNRRGLSQDILLNMCQIAKMS 834
Qy 841 YLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDEYHADGKVPKIMWALESI 900
Db 835 YLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDEYHADGKVPKIMWALESI 894
Qy 901 LRRRFTHQSDVMSYGVTVWELMTGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYMI 960
Db 895 LRRRFTHQSDVMSYGVTVWELMTGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYMI 954
Qy 961 MVKCMWIDSECRPRFRELVSFSEMRARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1020
Db 955 MVKCMWIDSECRPRFRELVSFSEMRARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1014
Qy 1021 GDLVDAEYLVPQOGFFCPDPAAGAGGMVHRHSSSTRSGGDLTLGLEPSEEEAPRSP 1080
Db 1015 GDLVDAEYLVPQOGFFCPDPAAGAGGMVHRHSSSTRSGGDLTLGLEPSEEEAPRSP 1074
Qy 1081 LAPSEGAGSDVDFDGLGMAAGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAPLTCS 1140
Db 1075 LAPSEGAGSDVDFDGLGMAAGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAPLTCS 1134
Qy 1141 POPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPKNGVVKDVPFAGGAVENP 1200
Db 1135 POPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPKNGVVKDVPFAGGAVENP 1194
Qy 1201 EYLTPQGAAPQHPHPPAFSPAFNLVYWDQDPPERGAPPSTFKGTPPTAENPEYGLDVP 1260
Db 1195 EYLTPQGAAPQHPHPPAFSPAFNLVYWDQDPPERGAPPSTFKGTPPTAENPEYGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255

RESULT 3
AAB60167
ID AAB60167 standard; Protein; 1255 AA.
AC AAB60167;
XX
XX
XX 03-APR-2001 (first entry)
DT HER2 transgene plasmid construct encoded protein.
DE
DE
XX Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
XX antibody.
XX Homo sapiens.
OS Synthetic.
XX
XX WO200100244-A2.
XX
XX 04-JAN-2001.
FD
XX
XX 23-JUN-2000; 2000WO-US17229.
XX
XX 25-JUN-1999; 99US-0141316.
PR 16-MAR-2000; 2000US-0189844.
XX
XX (GETH ) GENENTECH INC.
PA
XX Erickson,S, Schwall R;
PI
```

```
XX
DR WPI: 2001-061962/07.
DR N-PSDB; AAF24297.
XX
PT Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
PT conjugating the antibody to a maytansinoid -
XX
XX Example 3; Fig 4; 92pp; English.
XX
XX The present invention provides a method of treating cancer by
XX administering a conjugate of anti-ErbB antibody with a maytansinoid. In
XX particular, the antibody is directed against ErbB2 (also known as HER2
XX and p185neu). The method is particularly useful in the treatment of
XX breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
XX colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ Sequence 1255 AA;
Query Match 98.08; Score 6713; DB 22; Length 1255;
Best Local Similarity 98.58; Pred. No. 0;
Matches 1242; Conservative 4; Mismatches 9; Indels 6; Gaps 2;
Qy 1 MELAALCRWGLLLALLPPGAASQVCTGDMKRLPASPETHDMLRHLYGQCVVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASQVCTGDMKRLPASPETHDMLRHLYGQCVVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQFNFTVSVFWRVPKVSASHLEORLIRVRGTQLBEDNYAL 120
Db 61 ELTYLPTNASLSFLQDIQEVQ--GYVLIHNRQVPL----ORLIRVRGTQLBEDNYAL 114
Qy 121 AVLNDGDPNNTTPTVGTASPGGLRELQRLSITLTKGGVLIQRNPOLCYQDTILWKDIFH 180
Db 115 AVLNDGDPNNTTPTVGTASPGGLRELQRLSITLTKGGVLIQRNPOLCYQDTILWKDIFH 174
Qy 181 KNNQALTLTDNRSRACHPCSPMKGSRGWCWSESSEDCQSLTRTVCAAGCARCKGPLPTD 240
Db 175 KNNQALTLTDNRSRACHPCSPMKGSRGWCWSESSEDCQSLTRTVCAAGCARCKGPLPTD 234
Qy 241 CCHQCAAGCTGPKHSDCLACLPNHSIGICELHCPALVTYNTDTFESMPNPEGYRTFGAS 300
Db 235 CCHQCAAGCTGPKHSDCLACLPNHSIGICELHCPALVTYNTDTFESMPNPEGYRTFGAS 294
Qy 301 CVTACPNYLSLTDVGSCTLVCPLNHQRVTAEDGTQRCCKSPKPCARVCYGLGMHLEVR 360
Db 295 CVTACPNYLSLTDVGSCTLVCPLNHQRVTAEDGTQRCCKSPKPCARVCYGLGMHLEVR 354
Qy 361 AVTSANIQEFAGCKKIFGSLAFPLESPDGPASNTAPLOPEQLQVFTLEETITGLYISA 420
Db 355 AVTSANIQEFAGCKKIFGSLAFPLESPDGPASNTAPLOPEQLQVFTLEETITGLYISA 414
Qy 421 WPDSPDLVSFQNLQVIRGRILHNGAYSLTQGLIGISWLGRLSRLRELSGLALIHNNTHL 480
Db 415 WPDSPDLVSFQNLQVIRGRILHNGAYSLTQGLIGISWLGRLSRLRELSGLALIHNNTHL 474
Qy 481 CFVHTVPWDQLFRNPHOALLHTANRPEDECVGEGGLACHQLCARGHCHGPGTQCVCNCSQF 540
Db 475 CFVHTVPWDQLFRNPHOALLHTANRPEDECVGEGGLACHQLCARGHCHGPGTQCVCNCSQF 534
Qy 541 LRGOECVEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDDP 600
Db 535 LRGOECVEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDDP 594
Qy 601 FCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKCGCPAQRASPLTSI 660
Db 595 FCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKCGCPAQRASPLTSI 654
Qy 661 VSAVVGILLVVVLGVVFGILLIKRQOKIRKYVIMRLLQETELVPLTPSGAMPNQAOVRI 720
Db 655 VSAVVGILLVVVLGVVFGILLIKRQOKIRKYVIMRLLQETELVPLTPSGAMPNQAOVRI 714
Qy 721 LKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVM 780
```

Db 715 LKETELRKVKVLGSGAFQTVVKGWIPDGENVKIPVAIKVIRENTSPKANKEIIDAEAYVM 774
Qy 781 AGVGSPPVSRLLGICLTSTVQLTQMPYGCGLLDHVRNRRGLSGQDILLNWCMIQAKGMS 840
Db 775 AGVGSPPVSRLLGICLTSTVQLTQMPYGCGLLDHVRNRRGLSGQDILLNWCMIQAKGMS 834
Qy 841 YLEVDRLVHRDLAARNVLKSPNVHKITDFGLARLLDIDETEHADGGKVPDKWMALESI 900
Db 835 YLEVDRLVHRDLAARNVLKSPNVHKITDFGLARLLDIDETEHADGGKVPDKWMALESI 894
Qy 901 LRRRFTHQSDVWSYGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVYMI 960
Db 895 LRRRFTHQSDVWSYGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVYMI 954
Qy 961 MVKCMWIDSECRPRFRELVSFSEMRARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDM 1020
Db 955 MVKCMWIDSECRPRFRELVSFSEMRARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDM 1014
Qy 1021 GDLVDAEYLVPOQGFPCPDPAAGAGMVHRRSSSTRSGGDLTLGLEPSEBEAPRSP 1080
Db 1015 GDLVDAEYLVPOQGFPCPDPAAGAGMVHRRSSSTRSGGDLTLGLEPSEBEAPRSP 1074
Qy 1081 LAPSEGAGSDVFDGDLGMAAGKQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCS 1140
Db 1075 LAPSEGAGSDVFDGDLGMAAGKQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCS 1134
Qy 1141 PQPEYVNPQDVRPQPPREGPLPAARPAGATLBRAKTLPSPKNGVWVDVFAFGAVENP 1200
Db 1135 PQPEYVNPQDVRPQPPREGPLPAARPAGATLBRAKTLPSPKNGVWVDVFAFGAVENP 1194
Qy 1201 EYLTPQGAAPQPHPPAPFAPFONLYWQDPPERGAPPTFKGTPTAENPEYGLDVP 1260
Db 1195 EYLTPQGAAPQPHPPAPFAPFONLYWQDPPERGAPPTFKGTPTAENPEYGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255

RESULT 4
AAU74545
ID AAU74545 standard; Protein; 1255 AA.
XX AC AAU74545;
XX DT 23-APR-2002 (first entry)
XX DE Human HER2 (Erbb2) polypeptide.
XX Human; HER2; Erbb2; epidermal growth factor receptor; receptor;
KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
KW thyroid; pancreas; prostate; bladder; Erbb2; neuronal disorder;
KW glial disorder; astrocytal disorder; hypothalamic disorder;
KW glandular disorder; macrophagal disorder; epithelial disorder;
KW stromal disorder; blastocoealic disorder; inflammatory disorder;
KW angiogenic disorder; immunological disorder.
XX Homo sapiens.
XX OS
XX US2002001587-A1.
XX PN
XX 03-JAN-2002.
XX PD
XX 16-MAR-2001; 2001US-0811123.
XX PF
XX 16-MAR-2000; 2000US-189844P.
XX PR
XX 05-OCT-2000; 2000US-238327P.
XX PR
XX (ERIC/) ERICKSON S.
XX PA
XX (SCHW/) SCHWALL R.
XX PA
XX (SLIW/) SLIWKOWSKI M.

Erickson S, Schwall R, Sliwkowski M;

WPI; 2002-163686/21.

N-PSDB; ABK14058.

Treating tumour characterised by overexpression of epidermal growth factor receptor, ErbB or cancer in mammal, comprises administering anti-ErbB antibody-maytansinoid conjugate to the mammal -

Example 3; Fig 7; 93pp; English.

The invention relates to treating a tumour in a mammal, where the tumour is characterised by the overexpression of an epidermal growth factor receptor (ErbB) and does not respond or responds poorly, to treatment with an anti-ErbB antibody, comprising administering to the mammal an anti-ErbB antibody-maytansinoid conjugate. The method is useful for treating cancer or tumours of the breast, ovary, stomach, endometrium, salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas, prostate and bladder, preferably breast cancer. The breast cancer is a metastatic breast cancer or an aggressive form of metastatic breast cancer which overexpresses ErbB2. The method is also useful for treating neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal, epithelial, stromal, blastocoealic, inflammatory, angiogenic and immunological disorders. This sequence represents the human HER2 (Erbb2) polypeptide of the invention.

SQ Sequence 1255 AA;

Query Match 98.0%; Score 6713; DB 23; Length 1255;

Best Local Similarity 98.5%; Pred. No. 0;

Matches 1242; Conservative 4; Mismatches 9; Indels 6; Gaps 2;

Qy 1 MELAALCRWGLLALLPPGAASQVCTGTDKMLRASPETHLDMLRHLVGGCOVVOGNL 60

Db 1 MELAALCRWGLLALLPPGAASQVCTGTDKMLRASPETHLDMLRHLVGGCOVVOGNL 60

Qy 61 ELTYLPTNASLSFLQDIOEQVFNFTVSPWLRVPKVSASHLEQRLEIRVGTQLPEDNYAL 120

Db 61 ELTYLPTNASLSFLQDIOEQVFNFTVSPWLRVPKVSASHLEQRLEIRVGTQLPEDNYAL 114

Qy 121 AVLDNGDPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQNPQLCYODTILWKDIFH 180

Db 115 AVLDNGDPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQNPQLCYODTILWKDIFH 174

Qy 181 KNNQLALTLDITNRSRACHPCSPMKSGSRGWGSESSDCQSLTRTVACAGGCAKCKGLPTD 240

Db 175 KNNQLALTLDITNRSRACHPCSPMKSGSRGWGSESSDCQSLTRTVACAGGCAKCKGLPTD 234

Qy 241 CCEQCAAGCT3PKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 300

Db 235 CCEQCAAGCT3PKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 294

Qy 301 CVTACPNYLSLTDVGSCTLVCPLNHQBVTAEADGTQCEKCKSPCARVCYGLGMEHLREVR 360

Db 295 CVTACPNYLSLTDVGSCTLVCPLNHQBVTAEADGTQCEKCKSPCARVCYGLGMEHLREVR 354

Qy 361 AVTSANIOEPAGCKKIFGSLAFLPESDGPASNTAPLOPEQLQVFFETLEITGLYLISA 420

Db 355 AVTSANIOEPAGCKKIFGSLAFLPESDGPASNTAPLOPEQLQVFFETLEITGLYLISA 414

Qy 421 WPDLSPLDSVFNQLQVIRGRIHLNGAYSLTLOGIGISWLGRLSRLRELGSGLALIHNNTHL 480

Db 415 WPDLSPLDSVFNQLQVIRGRIHLNGAYSLTLOGIGISWLGRLSRLRELGSGLALIHNNTHL 474

Qy 481 CFVHTVPWDQLFRNPHQALLHTANRBEDECVGSLACHQLCARGHCWGPGPTQCVNCSQF 540

Db 475 CFVHTVPWDQLFRNPHQALLHTANRBEDECVGSLACHQLCARGHCWGPGPTQCVNCSQF 534

Qy 541 LRGOECVEECRLVQGLPREYVNHARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDDP 600

Db 535 LRGOECVEECRLVQGLPREYVNHARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDDP 594

Qy 601 FCVARCPGKVPDLSTYMPIWKFPDEEGACQPCPINCTHSCVDLDDDKGCPAQRASPLTSI 660

```
Db 595 FCVAPCSGVKFDLSYMEIPWKFPEDEGACQPCPINCSTHSCVDLDDKGCAPQASPLTSI 654
Qy 661 VSAVVGILLVVVLGVVFGILIKRQOKIRKYMRLLOSTELVEPLTPSGAMPNQAQWRI 720
Db 655 VSAVVGILLVVVLGVVFGILIKRQOKIRKYMRLLOSTELVEPLTPSGAMPNQAQWRI 714
Qy 721 LKETELRKVKVLGSGAGFTVVGKTIWIPDGENVKIPVAIKVRENTSPKANKEIIDEAYVM 780
Db 715 LKETELRKVKVLGSGAGFTVVGKTIWIPDGENVKIPVAIKVRENTSPKANKEIIDEAYVM 774
Qy 781 AGVGSPPVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRNRRGLSGQDILLNWCMIAGMS 840
Db 775 AGVGSPPVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRNRRGLSGQDILLNWCMIAGMS 834
Qy 841 YLEDVRLVHRDLAARNVLKSPNHWKITDFGLARLLDIDETEHADGKVPKIKWMALESI 900
Db 835 YLEDVRLVHRDLAARNVLKSPNHWKITDFGLARLLDIDETEHADGKVPKIKWMALESI 894
Qy 901 LRRRFTHQSDVWSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVYMI 960
Db 895 LRRRFTHQSDVWSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVYMI 954
Qy 961 MVKCMWIDSECRPRFRELVSFMRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDM 1020
Db 955 MVKCMWIDSECRPRFRELVSFMRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDM 1014
Qy 1021 GDLVDAEYLVPOQGFPCDPAPGAGGMVHRRHSSSTRSGGDLITLGLPSEBEAPRSP 1080
Db 1015 GDLVDAEYLVPOQGFPCDPAPGAGGMVHRRHSSSTRSGGDLITLGLPSEBEAPRSP 1074
Qy 1081 LAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPLORYSEDPVPLPSETDGYVAPLTCS 1140
Db 1075 LAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPLORYSEDPVPLPSETDGYVAPLTCS 1134
Qy 1141 POPEYVNPQDVRPOPSPREGPLPAARPAAGATLERAKTSLSPGKNGVVKDVFAGGAVENP 1200
Db 1135 POPEYVNPQDVRPOPSPREGPLPAARPAAGATLERAKTSLSPGKNGVVKDVFAGGAVENP 1194
Qy 1201 EYLTPQGAAPQPHPPAFSAPFNLYYWDQDPPERGAPPTFKGTPTAENPEYLGLDVP 1260
Db 1195 EYLTPQGAAPQPHPPAFSAPFNLYYWDQDPPERGAPPTFKGTPTAENPEYLGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255

RESULT 5
AAW01111
ID AAW01111 standard; Protein; 1255 AA.
XX AC AAW01111;
XX AC AAW01111;
DT 01-JAN-1997 (first entry)
XX DE HER-2/neu protein.
KW HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW breast cancer; ovary cancer; colon cancer; lung cancer;
KW prostate cancer; immunisation; tumour; vaccine; vector.
XX OS Homo sapiens.
XX Key Location/Qualifiers
FT Domain 676..1255
FT /label= intracellular domain
FT /note= "claimed domain, useful for immunisation"
XX PN W09630514-A1.
XX PD 03-OCT-1996.
XX
```

```
PF 28-MAR-1996; 96WO-US01689.
PR 31-MAR-1995; 95US-0414417.
XX (UNIW ) UNIV WASHINGTON.
XX Cheever MA, Disis ML;
XX MPI; 1996-455361/45.
XX N-PSDB; AAT40739.
XX DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
PT treatment of malignancies with which the HER-2/neu oncogene is
PT associated
XX
XX Claim 2; Page 56-61; 7lpp; English.
XX
XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.
XX
XX SQ Sequence 1255 AA;
```

```
Query Match 98.0%; Score 6707; DB 17; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1240; Conservative 5; Mismatches 10; Indels 6; Gaps 2;
Qy 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPTHLDMLRHLVGGCQVQGNL 60
Db 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPTHLDMLRHLVGGCQVQGNL 60
Qy 61 ELTYLPNALSFLQDIOEQVFNFTVSFWLRVPKVSASHLEORLIRVRGTQLPEDNYAL 120
Db 61 ELTYLPNALSFLQDIOEQV--GYVLIANQVRQVPL---QELRIRVGTQLPEDNYAL 114
Qy 121 AVLNDGDLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFH 180
Db 115 AVLNDGDLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFH 174
Qy 181 KNNQALTLIDTNRSRACHPCSPMCKSGRCWGSSEDCQSLTRTVACGGCARKGPLPTD 240
Db 175 KNNQALTLIDTNRSRACHPCSPMCKSGRCWGSSEDCQSLTRTVACGGCARKGPLPTD 234
Qy 241 CCHEOCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 300
Db 235 CCHEOCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 294
Qy 301 CVTACPNYLSLTDVGSCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 360
Db 295 CVTACPNYLSLTDVGSCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 354
Qy 361 AVTSANTIQEFAGCKKIFGSLAFIPESPDGPASNTAPLOPEQLOVFTLEITCYLYISA 420
Db 355 AVTSANTIQEFAGCKKIFGSLAFIPESPDGPASNTAPLOPEQLOVFTLEITCYLYISA 414
Qy 421 WPDSDLPSLVFQNLQVIRGRILHNGAYSLSLTQGLGISWLGRLSRLRELGSGLALHNNTHL 480
Db 415 WPDSDLPSLVFQNLQVIRGRILHNGAYSLSLTQGLGISWLGRLSRLRELGSGLALHNNTHL 474
Qy 481 CFVHTVPWDQLFRNPQOALLHTANRPEDECVGEGLAGCHQLCARGHCWGPQTQCVCNSQF 540
Db 475 CFVHTVPWDQLFRNPQOALLHTANRPEDECVGEGLAGCHQLCARGHCWGPQTQCVCNSQF 534
Qy 541 LRGOECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFSGPEADQCACAHYKDDP 600
Db 535 LRGOECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFSGPEADQCACAHYKDDP 594
```

```

Qy 601 FCVARCPSGVKPDLSYMPDIWKFDPBEGACQPCINCTHSCVDLDDKGCAPAQORASPLTSI 660
Db 595 FCVARCPSGVKPDLSYMPDIWKFDPBEGACQPCINCTHSCVDLDDKGCAPAQORASPLTSI 654
Qy 661 VSAVVGILLVVLGVVFGILLIKRQOKIRKTYTMRLLQETELVEPLTPSGAMPNOAQMRI 720
Db 655 ISAVVGILLVVLGVVFGILLIKRQOKIRKTYTMRLLQETELVEPLTPSGAMPNOAQMRI 714
Qy 721 LKETELRKVKVLSGAGFGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILLDEAYVM 780
Db 715 LKETELRKVKVLSGAGFGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILLDEAYVM 774
Qy 781 AGVSPYVSRLLGLCLSTVQLVLTQMLPYGCLLDHVRNRRGLSQDILLNCMOIAKMS 840
Db 775 AGVSPYVSRLLGLCLSTVQLVLTQMLPYGCLLDHVRNRRGLSQDILLNCMOIAKMS 834
Qy 841 YLEVDRLVHRDLAARNVLKSPNVKITDFGLARLLDIDETEHADGGKVPKMALES 900
Db 835 YLEVDRLVHRDLAARNVLKSPNVKITDFGLARLLDIDETEHADGGKVPKMALES 894
Qy 901 LRRRFTHQSDVMSYGVTVWELMTFCAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYMI 960
Db 895 LRRRFTHQSDVMSYGVTVWELMTFCAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYMI 954
Qy 961 MVKCMWIDSECRPRFRELVSFSEMRARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1020
Db 955 MVKCMWIDSECRPRFRELVSFSEMRARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1014
Qy 1021 GDLVDAEYLVPOQGFPCPDAPGAGGMVHRRSSSTRSGGDLTLGLEPSESEAPRSP 1080
Db 1015 GDLVDAEYLVPOQGFPCPDAPGAGGMVHRRSSSTRSGGDLTLGLEPSESEAPRSP 1074
Qy 1081 LAPSEGAGSDVFDGDLGMAKGLQSLPTHDPSPQLQRYSEDPVPLPSETDGYVAPLTCS 1140
Db 1075 LAPSEGAGSDVFDGDLGMAKGLQSLPTHDPSPQLQRYSEDPVPLPSETDGYVAPLTCS 1134
Qy 1141 POPEYVNPQDVRPQPPSPREGPLPAAPAGATLBRATKTLSPGKNGVVKDVPFAFGAVENP 1200
Db 1135 POPEYVNPQDVRPQPPSPREGPLPAAPAGATLBRATKTLSPGKNGVVKDVPFAFGAVENP 1194
Qy 1201 EYLTPOGGAAPQHPHPPAFSPAFNLVYWDODPPERGAPPSTFKGTPTAENPEYLGLDVP 1260
Db 1195 EYLTPOGGAAPQHPHPPAFSPAFNLVYWDODPPERGAPPSTFKGTPTAENPEYLGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255

RESULT 6
AAW92406
ID AAW92406 standard; Protein; 1255 AA.
XX AC AAW92406;
XX AC
XX DT 21-APR-1999 (first entry)
XX DE Human HER-2/neu oncogene protein.
XX KW HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
XX KW malignancy; treatment; tumour.
XX OS Homo sapiens.
XX FH
XX Key Location/Qualifiers
XX Region 676..1255
XX FT /note= "region which elicits immune response"
XX XX
XX PN US85869445-A.
XX FN
XX PD 09-FEB-1999.
XX XX
XX PF 01-APR-1996; 96US-0625101.

```

```

XX 01-APR-1996; 96US-0625101.
PR 17-MAR-1993; 93US-0033644.
PR 12-AUG-1993; 93US-0106112.
PR 31-MAR-1995; 95US-0414417.
XX (UNIW ) UNIV WASHINGTON.
XX Cheever MA, Disis ML;
XX MPI; 1999-152835/13.
DR N-PSDB; AAX01912.
XX
PT Use of HER-2/neu polypeptides - for eliciting an immune response to
PT an HER-2/neu associated malignancy, particularly for treating or
PT preventing tumours
XX
PS Claim 3; Column 31-38; 26pp; English.
XX
CC This sequence represents the human HER-2/neu oncogene protein. A fragment
CC of this protein is used in a method for eliciting or enhancing an immune
CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
CC B cells to produce an immune response to the HER-2/neu protein. The
CC method can be used for immunisation against a malignancy in which the
CC HER-2/neu oncogene is associated and in the treatment of an existing
CC tumour, or to prevent tumour occurrence or reoccurrence.
XX
SQ Sequence 1255 AA;
Query Match 98.0%; Score 6707; DB 20; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1240; Conservative 5; Mismatches 10; Indels 6; Gaps 2;
Qy 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPTHLDMLRHLYQGCVQVGNL 60
Db 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPTHLDMLRHLYQGCVQVGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQFNFTVSFWLRVPKVSASHLEQRLRVRCGTQLPEDNYAL 120
Db 61 ELTYLPTNASLSFLQDIQEVQ--GYVLIHNRVQVPL---QRLRVRCGTQLPEDNYAL 114
Qy 121 AVLDNGDPLNNTPTVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFH 180
Db 115 AVLDNGDPLNNTPTVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFH 174
Qy 181 KNNQALATLIDTNRSRACHPCSPMKGSRGCSSESDCSLTTRVCAGGCARCKGPLPTD 240
Db 175 KNNQALATLIDTNRSRACHPCSPMKGSRGCSSESDCSLTTRVCAGGCARCKGPLPTD 234
Qy 241 CHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 300
Db 235 CHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 294
Qy 301 CVTACPNYLSSTDVSGCTLVCPLNQEVTAEDGTQRCCKSPKPCARVCYGLGMEHLREVR 360
Db 295 CVTACPNYLSSTDVSGCTLVCPLNQEVTAEDGTQRCCKSPKPCARVCYGLGMEHLREVR 354
Qy 361 AVTSANIQEPAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEITGYLYISA 420
Db 355 AVTSANIQEPAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEITGYLYISA 414
Qy 421 WPDLSLPDLSVFNQLQVIRGRILHNGAYSLLTQGLGISWLGRLSRLRELGSGLALIHNNTHL 480
Db 415 WPDLSLPDLSVFNQLQVIRGRILHNGAYSLLTQGLGISWLGRLSRLRELGSGLALIHNNTHL 474
Qy 481 CFVHTVPWDLFRNPQALLHTANRDEBCVGBGLACHQLCARHGCWGPQPTQVNCVSQF 540
Db 475 CFVHTVPWDLFRNPQALLHTANRDEBCVGBGLACHQLCARHGCWGPQPTQVNCVSQF 534
Qy 541 LRQECVEECRVLQGLPREYVNAHCLPCHPEQOPNGSVTCFPGPADQCVACAHYKDDPP 600
Db 535 LRQECVEECRVLQGLPREYVNAHCLPCHPEQOPNGSVTCFPGPADQCVACAHYKDDPP 594

```

```
Qy 601 FCVARCPGVKPDLSYMPFIWKFPEDEGACQPCPINCSTHSCVDLDDKGCAPQASPLTSI 660
Db 595 FCVARCPGVKPDLSYMPFIWKFPEDEGACQPCPINCSTHSCVDLDDKGCAPQASPLTSI 654
Qy 661 VSAVGILLVVLGVVFGILIKRQOKIRKYTMRLLOETELVBPPLTPSGAMPNQAQWRI 720
Db 655 ISAVVGLLVVLGVVFGILIKRQOKIRKYTMRLLOETELVBPPLTPSGAMPNQAQWRI 714
Qy 721 LKETELRKVKVLGSGAGCTVYKGIWIPDGENVKIPVAIKVRENTSPKANKEIILDEAYVM 780
Db 715 LKETELRKVKVLGSGAGCTVYKGIWIPDGENVKIPVAIKVRENTSPKANKEIILDEAYVM 774
Qy 781 AGVSPVYSRLGLCLSTVOLTPQLMPYGCCLLDHVENRGRIGSQDILNWCMIAGKMS 840
Db 775 AGVSPVYSRLGLCLSTVOLTPQLMPYGCCLLDHVENRGRIGSQDILNWCMIAGKMS 834
Qy 841 YLEDVRLVHRDLAARNVLKSPNVHKITDFGLARLLDIDETEHADGGKVPKMWALESI 900
Db 835 YLEDVRLVHRDLAARNVLKSPNVHKITDFGLARLLDIDETEHADGGKVPKMWALESI 894
Qy 901 LRRRFTHQSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPORPDICTIDVYMI 960
Db 895 LRRRFTHQSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPORPDICTIDVYMI 954
Qy 961 MVKCMWIDSECRPFRELVSFSEMRADPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1020
Db 955 MVKCMWIDSECRPFRELVSFSEMRADPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1014
Qy 1021 GDLVDAEYLVPOQGFCCPDPAAGAGGMVHHRSSSTRSGGDLTLGLEPSEBEAPRSP 1080
Db 1015 GDLVDAEYLVPOQGFCCPDPAAGAGGMVHHRSSSTRSGGDLTLGLEPSEBEAPRSP 1074
Qy 1081 LAPSEGAGSDVFDGLGNAKGLQSLTPHDPSPLORYSEDPVPLPSETDGYVAPLTCS 1140
Db 1075 LAPSEGAGSDVFDGLGNAKGLQSLTPHDPSPLORYSEDPVPLPSETDGYVAPLTCS 1134
Qy 1141 POPEYVNOPDVRPOPSPREGPLPAARPAAGATLERAKTSLSPKNGVVKDVFAGGAVENP 1200
Db 1135 POPEYVNOPDVRPOPSPREGPLPAARPAAGATLERPKTSLSPKNGVVKDVFAGGAVENP 1194
Qy 1201 EYLTPQGGAAPOPHPPPAFSPAFDNLVYWDQDPPPERGAPPTFKGTPTAENPEYGLDVP 1260
Db 1195 EYLTPQGGAAPOPHPPPAFSPAFDNLVYWDQDPPPERGAPPTFKGTPTAENPEYGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255

RESULT 7
AAB21198
ID AAB21198 standard; protein; 1255 AA.
XX AC AAB21198;
XX DT 12-JAN-2001 (first entry)
XX DE Human HER-2/neu protein.
XX KW Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
XX KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
XX KW colon cancer.
XX OS Homo sapiens.
XX FN WO200044899-A1.
XX PD 03-AUG-2000.
XX PF 28-JAN-2000; 2000WO-US02164.
XX PR 29-JAN-1999; 99US-0117976.
XX
```

```
PA (CORI-) CORIXA CORP.
PA (SMIK ) SMITHKLINE BEECHAM.
PI Cheever MA, Gheysen D;
XX WPI; 2000-505976/45.
DR N-PSDB; AAA89736.
XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT useful for vaccinating against breast, ovarian, colon, lung and
PT prostate cancers -
XX Claim 52; Fig 7; 128pp; English.
XX The present sequence is the human HER-2/neu protein. It is a member of
CC the tyrosine kinase family of receptor-like glycoproteins and shows
CC homology to the epidermal growth factor receptor (EGFR). It probably
CC plays a part in cell growth and/or differentiation. The HER-2/neu
CC gene is an oncogene. An HER-2/neu fusion protein comprising a
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC domain may be used to treat or prevent cancer by eliciting or
CC enhancing an immune response to the HER-2/neu protein. It may be used
CC to treat malignancies such as breast, ovarian, colon, lung and
CC prostate cancers, and may be used as an antigen to vaccinate against
CC these neoplasias.
XX Sequence 1255 AA;
Qy Query Match 98.0%; Score 6707; DB 21; Length 1255;
Db Best Local Similarity 98.3%; Pred. No. 0;
Qy Matches 1240; Conservative 5; Mismatches 10; Indels 6; Gaps 2;
Db 1 MELAAALCRWGLLALLPPGAASQVCTGTDKMLRASPETHLDMLRHLXGQCVVQGNL 60
Qy 1 MELAAALCRWGLLALLPPGAASQVCTGTDKMLRASPETHLDMLRHLXGQCVVQGNL 60
Db 61 ELTYLPTNASLSFLQDIQEVQ--GYVLJAHNVQVPL---QRLRVRGTLQFEDNYAL 120
Qy 61 ELTYLPTNASLSFLQDIQEVQ--GYVLJAHNVQVPL---QRLRVRGTLQFEDNYAL 114
Db 121 AVLNDGPDPLNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFH 180
Qy 115 AVLNDGPDPLNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFH 174
Qy 181 KNNQALALTLDNRSRACHPCSPMCKSGRCWGSSEDCQSLTRTVACAGGACRCKGPLPTD 240
Db 175 KNNQALALTLDNRSRACHPCSPMCKSGRCWGSSEDCQSLTRTVACAGGACRCKGPLPTD 234
Qy 241 CCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 300
Db 235 CCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 294
Qy 301 CVTACPNYLSLTDVGSCTLVCPLNHQNVEATDGTQRCCKSPKPCARVCYGLGMHLEHVR 360
Db 295 CVTACPNYLSLTDVGSCTLVCPLNHQNVEATDGTQRCCKSPKPCARVCYGLGMHLEHVR 354
Qy 361 AVTSANTIOEFAGCKKIFGSLAFIPESFDGPASNTAPLOEQLOVFTLEETICVLYISA 420
Db 355 AVTSANTIOEFAGCKKIFGSLAFIPESFDGPASNTAPLOEQLOVFTLEETICVLYISA 414
Qy 421 WPDSLPLDSVFNQLQVIRGRILHNGAYSLSLTLQGLIGISWGLRSLRELGSGLALIHNNH 480
Db 415 WPDSLPLDSVFNQLQVIRGRILHNGAYSLSLTLQGLIGISWGLRSLRELGSGLALIHNNH 474
Qy 481 CFVHTVPWDQLFRNPHOALLHTANRPEDECVGEGACHQLCARGHCWGPGPTQCNCVSQF 540
Db 475 CFVHTVPWDQLFRNPHOALLHTANRPEDECVGEGACHQLCARGHCWGPGPTQCNCVSQF 534
Qy 541 LRGOECVVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDDPP 600
Db 535 LRGOECVVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDDPP 594
Qy 601 FCVARCPGVKPDLSYMPFIWKFPEDEGACQPCPINCSTHSCVDLDDKGCAPQASPLTSI 660
```


Db 595 FCVARGSGVDPDLSYMPWIKFPDEBACQPCPINCSTHSCVDLDDKQCPAQSRASPLTISI 654
Qy 661 VSADVGLLVVVLGVVFGILIKRQKQKIRKTYTMRLLQETELVEPLTPSGAMPNQAWRI 720
Db 655 ISADVGLLVVVLGVVFGILIKRQKQKIRKTYTMRLLQETELVEPLTPSGAMPNQAWRI 714
Qy 721 LKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEIIDEAYVM 780
Db 715 LKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEIIDEAYVM 774
Qy 781 AGVGSPPVSRLLGICLTSTVOLTPQMPYGCCLLDHVRNRCRLGSODLLNWCMIAGMS 840
Db 775 AGVGSPPVSRLLGICLTSTVOLTPQMPYGCCLLDHVRNRCRLGSODLLNWCMIAGMS 834
Qy 841 YLEDRVLRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNWALESI 900
Db 835 YLEDRVLRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNWALESI 894
Qy 901 LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICITIDVYMI 960
Db 895 LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICITIDVYMI 954
Qy 961 MVKCMWIDSECRPRFRELVSFSEWARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1020
Db 955 MVKCMWIDSECRPRFRELVSFSEWARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1014
Qy 1021 GDLVDAEYLVPOQGFPCPDPAAGGGMVHRHRSSTRSGGDLTLGLEPSEBEPASP 1080
Db 1015 GDLVDAEYLVPOQGFPCPDPAAGGGMVHRHRSSTRSGGDLTLGLEPSEBEPASP 1074
Qy 1081 LAPSEGAGSVDFDGLGMAAKGLQSLTPHDPSPLORYSEDPTVPLPSETDGYVAPLTCS 1140
Db 1075 LAPSEGAGSVDFDGLGMAAKGLQSLTPHDPSPLORYSEDPTVPLPSETDGYVAPLTCS 1134
Qy 1141 POPEYVNOQDVRPOPPSPREGPLPAAPAGATLERAKTLSPGKNGVVKDVFAGGAVENP 1200
Db 1135 POPEYVNOQDVRPOPPSPREGPLPAAPAGATLERAKTLSPGKNGVVKDVFAGGAVENP 1194
Qy 1201 EYLTPOGGAAPQHPHPPAFSPAFNLVYWDQDPPERGAPPSTFKGTPPTAENPEYLGLDVP 1260
Db 1195 EYLTPOGGAAPQHPHPPAFSPAFNLVYWDQDPPERGAPPSTFKGTPPTAENPEYLGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255

RESULT 8

AY84780
ID AY84780 standard; Protein; 1255 AA.

XX AC AY84780;

XX AC AY84780;

DT 08-AUG-2000 (first entry)

XX Amino acid sequence of the SPLICE erbb-2 receptor protein.

XX SPLICE erbb-2 receptor protein; cell transformation disorder; cancer;
XX tumor cell proliferation; tissue degeneration; arthropathy;
XX bone resorption; inflammatory disease; degenerative disorder;
XX wound healing.

XX Homo sapiens.

XX W0200020579-A1.

XX 13-APR-2000.

XX 01-OCT-1999; 99WO-CA00912.

XX 02-OCT-1998; 98US-0165192.

XX

(UYMC-) UNIV MCMASTER.

XX Muller WJ, Siegel PM;

XX WPI: 2000-303768/26.

XX N-PSDB; AAA14812.

XX Nucleic acid encoding an erbb 2 receptor protein designated SPLICE
XX erbb-2, inhibitors of the protein are useful for treatment of cancer -
XX Claim 3; Fig 2; 60pp; English.

XX The present sequence represents a SPLICE erbb-2 receptor protein. The
XX protein has an in-frame deletion of 16 amino acids, 2 of which are
XX conserved cysteine residues, compared to the unspliced protein. The
XX erbb-2 polynucleotide is used to construct probes for detecting
XX disorders of cell transformation such as cancer. Antibodies to the
XX protein may be used to detect SPLICE erbb-2 in a sample. Agents
XX (e.g. antisense oligonucleotides) which inhibit the expression of
XX SPLICE erbb-2 are useful for reducing tumor cell proliferation and
XX treating cancer. Substances which stimulate SPLICE erbb-2 are useful
XX for treating conditions involving damaged cells including conditions
XX in which degeneration of tissue occurs such as arthropathy, bone
XX resorption, inflammatory diseases, degenerative disorders of the
XX central nervous system and wound healing.

SQ Sequence 1255 AA;

Query Match 98.0%; Score 6707; DB 21; Length 1255;

Best Local Similarity 98.3%; Pred. No. 0;

Matches 1240; Conservative 5; Mismatches 10; Indels 6; Gaps 2;

Qy 1 MELAALCRWGLLALLPPGAASQVCTGTDMLRLPASPTHDLMLRHLYGQGVQGNL 60

Db 1 MELAALCRWGLLALLPPGAASQVCTGTDMLRLPASPTHDLMLRHLYGQGVQGNL 60

Qy 61 ELYLPTNASISFLQDIOEQVFNFTVSWLRVPKVSASHLEORLRIVRGTQLPEDNYAL 120

Db 61 ELYLPTNASISFLQDIOEQVFNFTVSWLRVPKVSASHLEORLRIVRGTQLPEDNYAL 114

Qy 121 AVLNDGDPNNTPVTGASPGGLRELQRLSLTELKGGVLIORNPOLCYQDTILWKDIFH 180

Db 115 AVLNDGDPNNTPVTGASPGGLRELQRLSLTELKGGVLIORNPOLCYQDTILWKDIFH 174

Qy 181 KNNQLALTLDITNRSRACHPCSPMKGSRGSCWGSSEDCQSLTRTVTCAGGCARCKGPLPTD 240

Db 175 KNNQLALTLDITNRSRACHPCSPMKGSRGSCWGSSEDCQSLTRTVTCAGGCARCKGPLPTD 234

Qy 241 CCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFSSMPNPEGRYTFGAS 300

Db 235 CCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFSSMPNPEGRYTFGAS 294

Qy 301 CVTACPNYLSLTDVGSCTLVCPLNHOEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 360

Db 295 CVTACPNYLSLTDVGSCTLVCPLNHOEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 354

Qy 361 AVTSANIQEPAGCKKIFGSLAFPEPSDGPASNTAPLOPEQLQVFTLSEITGLYLISA 420

Db 355 AVTSANIQEPAGCKKIFGSLAFPEPSDGPASNTAPLOPEQLQVFTLSEITGLYLISA 414

Qy 421 WPDLSLPSVFNQLQVIRGRILHNGAYSLSLQGLISWLGRLSLRELGSGLAIHHNTHL 480

Db 415 WPDLSLPSVFNQLQVIRGRILHNGAYSLSLQGLISWLGRLSLRELGSGLAIHHNTHL 474

Qy 481 CFVHTVPWDQLFRNPHQALLHTANRDECEVCGSLACHQLCARGHCWGPPTQCVNCSQF 540

Db 475 CFVHTVPWDQLFRNPHQALLHTANRDECEVCGSLACHQLCARGHCWGPPTQCVNCSQF 534

Qy 541 LRQECVCECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQCVACAHYKDDP 600

Db 535 LRQECVCECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQCVACAHYKDDP 594

Qy 601 FCVARCPGSKPDLSYMPIWKFPDEBACQPCPINCSTHSCVDLDDKGCAPAEQASPLTISI 660

Db 475 CFVHTVPMDQLFRNPQHALLTANRDECECVGEGLACHQLCARGHCWGPGPTQCVCNCSQF 534
Qy 541 LRQGECEVEECVILQGLPREYNARHCLPCHPECPONGSVTCFGEADQCACAHYKDP 600
Db 535 LRQGECEVEECVILQGLPREYNARHCLPCHPECPONGSVTCFGEADQCACAHYKDP 594
Qy 601 FCVARCPGVPDLISYMPDWKPFDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSI 660
Db 595 FCVARCPGVPDLISYMPDWKPFDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSI 654
Qy 661 VSAVVGILLVVLGVVFGILIKRQKIRKYTMRLLOETELVPLTPSGAMPNQAQMRI 720
Db 655 ISAVVGILLVVLGVVFGILIKRQKIRKYTMRLLOETELVPLTPSGAMPNQAQMRI 714
Qy 721 LKETELRKVKLGSGAFGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVM 780
Db 715 LKETELRKVKLGSGAFGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVM 774
Qy 781 AGVSPVYSRLIGLCTSTVOLVTQMPYGCLLDHVRENRLGSDLLNMCQIAKMS 840
Db 775 AGVSPVYSRLIGLCTSTVOLVTQMPYGCLLDHVRENRLGSDLLNMCQIAKMS 834
Qy 841 YLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALESI 900
Db 835 YLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALESI 894
Qy 901 LRRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMI 960
Db 895 LRRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMI 954
Qy 961 MVKCMIDSECRPRELVSFSRWARDPQRFVVIQNEEDLGPASPLDSTFYRSLLEDDDM 1020
Db 955 MVKCMIDSECRPRELVSFSRWARDPQRFVVIQNEEDLGPASPLDSTFYRSLLEDDDM 1014
Qy 1021 GDLVDAEYLVPOQGFPCDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEEAPRSP 1080
Db 1015 GDLVDAEYLVPOQGFPCDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEEAPRSP 1074
Qy 1081 LAPSEGAGSDVFDGLNGAAKGLQSLTHDPSPLORYSEDPVPLPSETDGYVAPLTC 1140
Db 1075 LAPSEGAGSDVFDGLNGAAKGLQSLTHDPSPLORYSEDPVPLPSETDGYVAPLTC 1134
Qy 1141 POPEVYNQDVRPQPPSPREGPLPAARPAAGATLERAKTLPKNGVWVDVAFGAVENP 1200
Db 1135 POPEVYNQDVRPQPPSPREGPLPAARPAAGATLERPKTLPKNGVWVDVAFGAVENP 1194
Qy 1201 EYLTPQGGAAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYGLDVP 1260
Db 1195 EYLTPQGGAAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255

RESULT 11
AAE24067 standard; Protein; 1255 AA.
XX ID AAE24067
XX AC AAE24067;
XX DT 23-SEP-2002 (first entry)
XX DE Human Her-2 protein.
XX KW Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KW hyperproliferative disorder; prophylaxis; inflammation; antisense;
KW tumour; gene therapy; phosphorothioate backbone.
XX OS Homo sapiens.
XX PN WO200222636-A1.
XX

PD 21-MAR-2002.
XX
PF 12-SEP-2001; 2001WO-US28572.
XX
PR 15-SEP-2000; 2000US-0663834.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Bennett CF, Cowse LM;
XX
DR WPI; 2002-471192/50.
DR N-PSDB; AAD38904.
XX
PT Novel antisense oligonucleotide which modulates the expression of Human
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT Inflammation or to prevent infection in humans -
XX
PS Example 13; Page 95-107; 116pp; English.
XX
CC The invention relates to antisense compounds targetted to a nucleic
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC that specifically hybridises with and inhibits the expression of Her2.
CC Antisense compounds of the invention are used for treating diseases or
CC conditions associated with Her2 such as hyperproliferative disorders
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC neural or cardiac cancer. They are also useful prophylactically e.g.
CC to prevent or delay infection, inflammation and tumour formation. The
CC invention is also used in gene therapy. The present sequence is human
CC Her-2 protein.
XX
SQ Sequence 1255 AA;
Query Match 98.08; Score 6707; DB 23; Length 1255;
Best Local Similarity 98.34; Pred. No. 0;
Matches 1240; Conservative 5; Mismatches 10; Indels 6; Gaps 2;
Qy 1 MELAALCRWGLLALLPPGAASCTQCTGDMKRLPASPETHLMLRHLVGGCQVQGNL 60
Db 1 MELAALCRWGLLALLPPGAASCTQCTGDMKRLPASPETHLMLRHLVGGCQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVQFNFTVSFWLRVPKVSASHLEQRLRIVRGTLQFEDNYAL 120
Db 61 ELTYLPTNASLSFLQDIOEVQ--GYVLIAHQVRQVPL---QRLRIVRGTLQFEDNYAL 114
Qy 121 AVLDNGPPLNTPVTGASPGGLRELOLSITELKGGVLIQRPOLCYQDTILWKDIFH 180
Db 115 AVLDNGPPLNTPVTGASPGGLRELOLSITELKGGVLIQRPOLCYQDTILWKDIFH 174
Qy 181 KNNQALTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVAGGCARCKGPLPTD 240
Db 175 KNNQALTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVAGGCARCKGPLPTD 234
Qy 241 CCHQCAAGCTGPKHSDCLACHFNHSGI CELHCPALVTYNTDTFESMPNPEGRTYTFGAS 300
Db 235 CCHQCAAGCTGPKHSDCLACHFNHSGI CELHCPALVTYNTDTFESMPNPEGRTYTFGAS 294
Qy 301 CVTACPNYILSTDVGSCTLYCPLHNQVTAEDGTQCEKCKPCARCYCYGLGMEHLREVR 360
Db 295 CVTACPNYILSTDVGSCTLYCPLHNQVTAEDGTQCEKCKPCARCYCYGLGMEHLREVR 354
Qy 361 AVTSANTQEFAGCKKIFGSLAFIPESFDGPASNTAPLQPEQLQVFTLEITGYLYISA 420
Db 355 AVTSANTQEFAGCKKIFGSLAFIPESFDGPASNTAPLQPEQLQVFTLEITGYLYISA 414
Qy 421 WPDLSPLDSVFNQLQVIRGRILHNGAYSFLTQGLGISWGLRSLRSLRELSGLALIHNNTHL 480
Db 415 WPDLSPLDSVFNQLQVIRGRILHNGAYSFLTQGLGISWGLRSLRSLRELSGLALIHNNTHL 474
Qy 481 CFVHTVPMDQLFRNPQHALLTANRDECECVGEGLACHQLCARGHCWGPGPTQCVCNCSQF 540
Db 475 CFVHTVPMDQLFRNPQHALLTANRDECECVGEGLACHQLCARGHCWGPGPTQCVCNCSQF 534
Qy 541 LRQGECEVEECVILQGLPREYNARHCLPCHPECPONGSVTCFGEADQCACAHYKDP 600

Db 535 LRQGCVEBECRLVQLGLPREYNARHCLPCHPEQCPQNGSVTCFGEADQCACAHYKDDP 594
Qy 601 FCVARCPSGVKPDLSYMPIWKFPDEGACQPCPCINCHTSCVDLDDKQCPAQRASPLTSI 660
Db 595 FCVARCPSGVKPDLSYMPIWKFPDEGACQPCPCINCHTSCVDLDDKQCPAQRASPLTSI 654
Qy 661 VSAVVGILLVVLGVVFGILIKRQOKIRKYTMRLQLQETELVEPLTPSGAMPNQAWRI 720
Db 655 ISAVVGILLVVLGVVFGILIKRQOKIRKYTMRLQLQETELVEPLTPSGAMPNQAWRI 714
Qy 721 LKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEIILDEAYVM 780
Db 715 LKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEIILDEAYVM 774
Qy 781 AGVGSPPVSRLLGICLTSTVOLTPMYPGCLLDHVRNRRGLSGQDNLNMCQIAKMS 840
Db 775 AGVGSPPVSRLLGICLTSTVOLTPMYPGCLLDHVRNRRGLSGQDNLNMCQIAKMS 834
Qy 841 YLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDDIDETEHADGKVPKMALES 900
Db 835 YLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDDIDETEHADGKVPKMALES 894
Qy 901 LRRRPTHOSDVWSYGVTVWELMTGAKYDGIIPAREIPDLLEKGERLPQPICTIDVYMI 960
Db 895 LRRRPTHOSDVWSYGVTVWELMTGAKYDGIIPAREIPDLLEKGERLPQPICTIDVYMI 954
Qy 961 MVKCMWIDSECRPRFRELVSFMRARDPQRFVVIQNEEDLGPASPLDSTFYRSLLEDDDM 1020
Db 955 MVKCMWIDSECRPRFRELVSFMRARDPQRFVVIQNEEDLGPASPLDSTFYRSLLEDDDM 1014
Qy 1021 GDLVDAEYLVPOQGFCCPDPAAGAGMVHRRSSSTRSGGDLTLGLEPSEEAAPRSP 1080
Db 1015 GDLVDAEYLVPOQGFCCPDPAAGAGMVHRRSSSTRSGGDLTLGLEPSEEAAPRSP 1074
Qy 1081 LAPSEGAGSDVFDGLGNAKGLQSLTHDPSPLQRYSEPTVPLSETGYVAPLTCS 1140
Db 1075 LAPSEGAGSDVFDGLGNAKGLQSLTHDPSPLQRYSEPTVPLSETGYVAPLTCS 1134
Qy 1141 PQPEYVNOQDVRPQPPSPREGPLPAARPAAGATLERAKTLPQKNGVVKDVFAGGAVENP 1200
Db 1135 PQPEYVNOQDVRPQPPSPREGPLPAARPAAGATLERAKTLPQKNGVVKDVFAGGAVENP 1194
Qy 1201 EYLTPQGAAPQPHPPAFSPAFNLVYWDQDPPERGAPPSTFKGTPPTAENPEYLGLDVP 1260
Db 1195 EYLTPQGAAPQPHPPAFSPAFNLVYWDQDPPERGAPPSTFKGTPPTAENPEYLGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255

RESULT 12
AAE20479
ID AAE20479 standard; Protein; 1255 AA.
XX
AC AAE20479;
XX
DT 01-JUL-2002 (first entry)
XX
DE Human Her-2/neu protein.
XX
KW Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
KW human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX
OS Homo sapiens.
XX
FH Key
FT Region
FT 1021..1030
FT /note= "Naturally processed HLA-B44-restricted epitope"
XX
PN W0200214503-A2.
XX

PD 21-FEB-2002.
XX 14-AUG-2001; 2001WO-US41733.
XX 14-AUG-2000; 2000US-235152P.
PR 28-SEP-2000; 2000US-236428P.
PR 21-FEB-2001; 2001US-270520P.
XX (CORI-) CORIXA CORP.
XX Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
PI McNeill PD, Vedvick TS;
XX WPI; 2002-280758/32.
DR N-PSDB; AAD32743.
XX
PT Novel isolated Her-2/Neu polypeptide composition useful for therapy,
PT prevention and diagnosis of cancer, preferably breast cancer -
XX
PS Disclosure; Page 114-117; 129pp; English.
XX
CC The invention relates to an isolated Her-2/Neu polypeptide composition
CC effective for eliciting an immune response. The invention is useful for
CC eliciting an immune response in a patient, where the patient is human
CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
CC The composition is useful for the therapy and diagnosis of cancer,
CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
CC and other compositions for the diagnosis, prevention and treatment of
CC human malignancies for stimulating and/or expanding T cells specific for
CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a
CC patient. The invention is useful for stimulating a T cell response in a
CC human patient, as probe or primer for nucleic acid hybridisation, to
CC selectively form duplex molecules with complementary stretches of the
CC entire Her-2/Neu gene or gene fragments of interest to isolate a full
CC length gene from a suitable library, and to direct expression of a
CC polypeptide in appropriate host cells. The composition is useful in
CC prophylactic or therapeutic applications and for the treatment of cancer,
CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-
CC associated malignancies. The invention is useful in gene therapy. The
CC present sequence is human Her-2/neu protein.
XX
SQ Sequence 1255 AA;

Query Match 98.0%; Score 6707; DB 23; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1240; Conservative 5; Mismatches 10; Indels 6; Gaps 2;
Qy 1 MEAALCRWGLLLALLPPGAASQVCTGDMKRLPASPETHLDMLRHLYGCGVQGNL 60
Db 1 MEAALCRWGLLLALLPPGAASQVCTGDMKRLPASPETHLDMLRHLYGCGVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEQVFNFTVSFWLRVPKVSASHLEORLVRGTOLPEDNYAL 120
Db 61 ELTYLPTNASLSFLQDIOEQVFNFTVSFWLRVPKVSASHLEORLVRGTOLPEDNYAL 114
Qy 121 AVLNDGDLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFH 180
Db 115 AVLNDGDLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFH 174
Qy 181 KNNQALATLIDTNRSRACHPCSPMKGSRGWSGSSDCQSLTRTVACGACRCKGPIPTD 240
Db 175 KNNQALATLIDTNRSRACHPCSPMKGSRGWSGSSDCQSLTRTVACGACRCKGPIPTD 234
Qy 241 CCHQCAAGCTGPKHSDCLACILHFNHSGICELHCPALVTYNTDTFFSMNPGEYRTFGAS 300
Db 235 CCHQCAAGCTGPKHSDCLACILHFNHSGICELHCPALVTYNTDTFFSMNPGEYRTFGAS 294
Qy 301 CVTACPNYLSTDVGSCTLLVCPLNHNOEVTADGTQRCCKSKPCARVCYGLGHEHLREVR 360
Db 295 CVTACPNYLSTDVGSCTLLVCPLNHNOEVTADGTQRCCKSKPCARVCYGLGHEHLREVR 354
Qy 361 AVTSANIQEFAGCKKIFGSLAFIPESFGDPASNTAPLOPELOQVFTETLEETGYLYISA 420
Db 361 AVTSANIQEFAGCKKIFGSLAFIPESFGDPASNTAPLOPELOQVFTETLEETGYLYISA 420

Db 355 AVTSANIQFAGCKKIFGSLAPFESFDGDPASNTAPLOPQLOVFETLEBITGYLVISA 414
Qy 421 WPSLPLDSVFQNLQVIRGRILHNGAYSLTIQGLIGISWLGRLSRRELGSGLALIHNNTHL 480
Db 415 WPSLPLDSVFQNLQVIRGRILHNGAYSLTIQGLIGISWLGRLSRRELGSGLALIHNNTHL 474
Qy 481 CFVHTVPMWDLFRPNPHQALLHTANRPEDECVEGEGCHOLCARGHCWGPGTQCVCNSQF 540
Db 475 CFVHTVPMWDLFRPNPHQALLHTANRPEDECVEGEGCHOLCARGHCWGPGTQCVCNSQF 534
Qy 541 LRGOECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDP 600
Db 535 LRGOECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDP 594
Qy 601 FCVARCPGKVPDLSYMPDIWPFDEGACQPCPINTHSCVDLDDKGCAPAEORASPLTSI 660
Db 595 FCVARCPGKVPDLSYMPDIWPFDEGACQPCPINTHSCVDLDDKGCAPAEORASPLTSI 654
Qy 661 VSUVGILLVVVLGVVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAOWRI 720
Db 655 ISAVVGILLVVVLGVVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAOWRI 714
Qy 721 LKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEIILDEAYVM 780
Db 715 LKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEIILDEAYVM 774
Qy 781 AGVSPYVSRLLGLTSTVOLVTLQMPYGLLDHVRNRRGLSGQDLNNCMQIAKMS 840
Db 775 AGVSPYVSRLLGLTSTVOLVTLQMPYGLLDHVRNRRGLSGQDLNNCMQIAKMS 834
Qy 841 YLEDVRLVHRDLAARNVLKSPNVHKITDFGLARLLDIDETEHADGKVPKKNWALESI 900
Db 835 YLEDVRLVHRDLAARNVLKSPNVHKITDFGLARLLDIDETEHADGKVPKKNWALESI 894
Qy 901 LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMI 960
Db 895 LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMI 954
Qy 961 MVKWMIDSECRPRELVSFESRMARDPQRFVVIQNEGLGASPLDSTFYRSLLDDDM 1020
Db 955 MVKWMIDSECRPRELVSFESRMARDPQRFVVIQNEGLGASPLDSTFYRSLLDDDM 1014
Qy 1021 GDLVDAEYLVPOQGFCCPDAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEAAPRSP 1080
Db 1015 GDLVDAEYLVPOQGFCCPDAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEAAPRSP 1074
Qy 1081 LAPSEGAGSDVFDGLGMAAGKLGSLQSLPFDPSPLQRYSEDPVPLPSETDGYVAPLTCS 1140
Db 1075 LAPSEGAGSDVFDGLGMAAGKLGSLQSLPFDPSPLQRYSEDPVPLPSETDGYVAPLTCS 1134
Qy 1141 POPEYVQPDVVRPQPPSPREGPLPAARAGATLERAKTILSPQKNGVVKDVFAGGAVENP 1200
Db 1135 POPEYVQPDVVRPQPPSPREGPLPAARAGATLERAKTILSPQKNGVVKDVFAGGAVENP 1194
Qy 1201 EYLTPQGAAPQHPHPPAFSADFNLVYWDQPPERGAAPPSTFKTGTPTAENPEYVGLDVP 1260
Db 1195 EYLTPQGAAPQHPHPPAFSADFNLVYWDQPPERGAAPPSTFKTGTPTAENPEYVGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255

RESULT 13

AA051143
ID AA051143 standard; Protein; 1255 AA.

XX
AC AA051143;

XX 17-JUN-2002 (first entry)

XX Human Her-2/neu oncogene-encoded p185 glycoprotein.

KW Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
tyrosine kinase; receptor; c-erbB2; gene therapy.

XX Homo sapiens.

Key Location/Qualifiers
Domain 1..653
/note= "extracellular domain"
Domain 676..1255
/note= "intracellular domain"
Domain 990..1255
/note= "phosphorylation domain"

PN WO200212341-A2.

XX 14-FEB-2002.

XX 03-AUG-2001; 2001WO-US24283.

XX 03-AUG-2000; 2000US-0632507.

XX (CORI-) CORIXA CORP.
XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Cheever MA, Gheysen D;

XX WPI; 2002-241743/29.

XX N-PSDB; ABA92250.

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting
or enhancing an immune response to the protein, has Her-2/neu
extracellular domain fused to Her-2/neu intracellular or
phosphorylation domain

XX Claim 68; Fig 7; 141pp; English.

XX The present sequence is that of human Her-2/neu (p185 glycoprotein
or c-erbB2). An oncogenic self-protein and target for anti-cancer
vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
in a variety of cancers, including breast, ovarian, colon, lung and
prostate cancer. Her-2/neu is a member of the tyrosine kinase
family of receptor-like glycoproteins. It comprises an extracellular
domain with homology to the epidermal growth factor receptor
(EGFR), a highly hydrophobic transmembrane domain and a C-terminal
intracellular domain that also shows homology to EGFR. Its
overexpression correlates with a poor prognosis in breast and
ovarian cancers. The invention provides Her-2/neu fusion
proteins, nucleic acids encoding them, viral vectors, and vaccines
comprising the fusion proteins or nucleic acid molecules. In
preferred fusion proteins, the extracellular domain of a Her-2/neu
protein is fused to a Her-2/neu intracellular domain or
phosphorylation domain (or its beta-fragment). An immune
response to Her-2/neu protein is elicited or enhanced by
administering the fusion protein in the form of a vaccine, or by
transfecting cells of an animal *ex vivo* with a nucleic acid
encoding the fusion protein, and delivering the transfected cells
to the animal. The fusion proteins, nucleic acids, and isolated
specific T-cells are useful for inhibiting the development of a
cancer, especially breast, ovarian, colon, lung or prostate cancer
in a patient. T cells that specifically react with a Her-2/neu
fusion protein can be used to remove tumour cells from a sample in
order to inhibit the development of cancer in a patient.

XX Sequence 1255 AA;

Query Match 98.0%; Score 6707; DB 23; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1240; Conservative 5; Mismatches 10; Indels 6; Gaps 2;

Qy 1 MELAALCRWGLLLALLPFGAASCTGTDKMLRPLASPETHLDMRLHYGCGVQGNL 60

Db 1 MELAALCRWGLLLALLPFGAASCTGTDKMLRPLASPETHLDMRLHYGCGVQGNL 60


```
Db 1135 PQEYVNPQDVRPQPSPREGPLPAARAGATLERPKTSLSPGKNGVVKDVFAGGAVENP 1194
Qy 1201 EYLTPOGGAAPQHPHPAPSPAFDNLYYWDODPERGAPBSTFKGTPTAENPEYLGLDVP 1260
Db 1195 EYLTPOGGAAPQHPHPAPSPAFDNLYYWDODPERGAPBSTFKGTPTAENPEYLGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255

RESULT 15
AAR39568
ID AAR39568 standard; Protein; 1433 AA.
XX AAR39568;
XX 07-FEB-1994 (first entry)
XX Sequence of c-erbB-2 tumour antigen.
XX Tumour antigen; c-erbB-2; glycoprotein.
XX Homo sapiens.
XX WO9316185-A.
XX 19-AUG-1993.
XX 05-FEB-1993; 93WO-US01055.
XX 06-FEB-1992; 92US-0831967.
XX (CETU) CETUS ONCOLOGY CORP.
XX (CREA-) CREATIVE BIOMOLECULES INC.
XX Houston LL, Huston JS, Oppermann H, Ring DB;
XX WPI; 1993-272889/34.
XX N-PSDB; AAQ46083.
XX New single chain Fv polypeptide binding to C-erbB-2 tumour
XX antigen - for imaging or treating breast or ovarian cancer etc.
XX Disclosure; pages 48-54; 87pp; English.
XX c-erbB-2 refers to a protein antigen expressed on the surface of
XX tumour cells, such as breast and ovarian tumour cells, which is an
XX approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
XX pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents
XX the location of a stop codon in AAQ46083.
XX Sequence 1433 AA;
SQ Query Match 97.3%; Score 6664; DB 14; Length 1433;
Best Local Similarity 97.8%; Pred. No. 0;
Matches 1233; Conservative 7; Mismatches 15; Indels 6; Gaps 2;

Qy 1 MELAALCRWGLLALLPFGAASQVCTGTDMKRLPASPTHLDMLRHLVQGCQVQGNL 60
Db 1 MELAALCRWGLLALLPFGAASQVCTGTDMKRLPASPTHLDMLRHLVQGCQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIEQVQFNFTVSFWLRVPKVSASHLEQRLRVRGTLQFEDNYAL 120
Db 61 ELTYLPTNASLSFLQDIEQVQFNFTVSFWLRVPKVSASHLEQRLRVRGTLQFEDNYAL 114
Qy 121 AVLNDGDPPLNNTTPTVTCASPGGLRELQRLSITELTKGGLVLIQNPQLCYQDTILWKDIFH 180
Db 115 AVLNDGDPPLNNTTPTVTCASPGGLRELQRLSITELTKGGLVLIQNPQLCYQDTILWKDIFH 174
Qy 181 KNNQALTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVACAGGCARCKGPLPTD 240
Db 175 KNNQALTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVACAGGCARCKGPLPTD 234
```

```
Db 61 ELTYLPTNASLSFLQDIEQVQFNFTVSFWLRVPKVSASHLEQRLRVRGTLQFEDNYAL 114
Qy 121 AVLNDGDPPLNNTTPTVTCASPGGLRELQRLSITELTKGGLVLIQNPQLCYQDTILWKDIFH 180
Db 115 AVLNDGDPPLNNTTPTVTCASPGGLRELQRLSITELTKGGLVLIQNPQLCYQDTILWKDIFH 174
Qy 181 KNNQALTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVACAGGCARCKGPLPTD 240
Db 175 KNNQALTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVACAGGCARCKGPLPTD 234
Qy 241 CCHQCAAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDFESMPNPEGRYTFGAS 300
Db 235 CCHQCAAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDFESMPNPEGRYTFGAS 294
Qy 301 CVTACPNYLTSTVGSCCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 360
Db 295 CVTACPNYLTSTVGSCCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 354
Qy 361 AVTSANTOEFAGCKKIFGSLAFLESFDPGDPASNTAPLOEQLOVFTEITGVLVISA 420
Db 355 AVTSANTOEFAGCKKIFGSLAFLESFDPGDPASNTAPLOEQLOVFTEITGVLVISA 414
Qy 421 WPDLSLPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRELGSGLAIHNNTHL 480
Db 415 WPDLSLPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRELGSGLAIHNNTHL 474
Qy 481 CFVHTVPMQDLFRNPHQALLTANRPEDECVGEGGLACHQLCARGHCWGPGTQCVCNSQF 540
Db 475 CFVHTVPMQDLFRNPHQALLTANRPEDECVGEGGLACHQLCARGHCWGPGTQCVCNSQF 534
Qy 541 LRQCEVCECVLQGLPREYNARHCLPCHPECOPQNGSVTCFGEADQCVACAHYKDDP 600
Db 535 LRQCEVCECVLQGLPREYNARHCLPCHPECOPQNGSVTCFGEADQCVACAHYKDDP 594
Qy 601 FCVARCPGVPDLSPYMWKFPDEGACQPCPINCSTHSCVDLDDKGPAPQASPLTSI 660
Db 595 FCVARCPGVPDLSPYMWKFPDEGACQPCPINCSTHSCVDLDDKGPAPQASPLTSI 654
Qy 720 VSAVVGILLVVLGVVFGILIKRQOKIRKYMTRRLLOETELVPLPSPGAMPNQAOMRI 720
Db 714 VSAVVGILLVVLGVVFGILIKRQOKIRKYMTRRLLOETELVPLPSPGAMPNQAOMRI 714
Qy 781 LKETELRKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVRENTSPKANKEILDEAYVM 780
Db 774 LKETELRKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVRENTSPKANKEILDEAYVM 774
Qy 840 AGVSPYVSRLLIGLICTSTVOLVQLMPYGCILLDHVRENRLGSGQDILLNMCQIAGMS 840
Db 834 AGVSPYVSRLLIGLICTSTVOLVQLMPYGCILLDHVRENRLGSGQDILLNMCQIAGMS 834
Qy 900 YLEVDVLRVHRLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMWALESI 900
Db 894 YLEVDVLRVHRLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMWALESI 894
Qy 960 LRRRPTHOSDVSXGVTVWELMTCAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMYI 960
Db 954 LRRRPTHOSDVSXGVTVWELMTCAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMYI 954
Qy 1020 MVKCMWIDSECRPRRELVSFSSRMARDPQRFVVIQNEEDLGPASPLDSTFYRSLLEDDDM 1020
Db 1014 MVKCMWIDSECRPRRELVSFSSRMARDPQRFVVIQNEEDLGPASPLDSTFYRSLLEDDDM 1014
Qy 1080 GDLVDAEYLVPOQGFCCPDAPAGAGMVHHRHSSSTRSGGDLITLGLPSESEAPRSP 1080
Db 1074 GDLVDAEYLVPOQGFCCPDAPAGAGMVHHRHSSSTRSGGDLITLGLPSESEAPRSP 1074
Qy 1140 LAPSEGAGSDVFDGLGMAKGLQSLPDPSPLOQYSEDPVPLPSETDGYVAPLTC 1140
Db 1134 LAPSEGAGSDVFDGLGMAKGLQSLPDPSPLOQYSEDPVPLPSETDGYVAPLTC 1134
Qy 1200 PQEYVNPQDVRPQPSPREGPLPAARAGATLERAKTSLSPGKNGVVKDVFAGGAVENP 1200
```


Search completed: July 22, 2003, 08:40:22
Job time : 43.1589 secs

QY	241	CCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTPESMNPDEGRYTFGAS	300
Db	235		
QY	301	CVTACPNYLVSTDVGSCTLVCPHNOEVTAEQGTORCEKCKPCARVCYGLGMEHLREVR	360
Db	295		
QY	361	AVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEBITGYLYISA	420
Db	355		
QY	421	WPSLPLDLSVFQNLQVIRGRILHNGAYSILTLQGLGISWLGRLSRRELGSGLALIHNNTHL	480
Db	415		
QY	481	CFVHTVPWDQLFRNPQHALLHTANRDECECVGEGACHQLCARGHCWGPGTQCVCNSQF	540
Db	475		
QY	541	LRGQECVEECRVLQGLPREYVNAHRLCPCHPECPQNGSVTCFGEADQCVACAHYKOPP	600
Db	535		
QY	601	FCVARCPGVPDLSYMPIWKPDEEGACQPCPCINCTHSCVDLDDKGCAPAQRASPLTSI	660
Db	595		
QY	661	VSAVVGILLVVVLGVVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNQAQMRI	720
Db	655		
QY	721	LKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILLDEAYVM	780
Db	715		
QY	781	AGVGSPPVSRLLGICLTSTVLTQMPYGCCLDHVRENRLGSDQLLNWCMIAGMS	840
Db	775		
QY	841	YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLIDIDETEHADGGKVPKWMALESI	900
Db	835		
QY	901	LRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPPICTIDVYMI	960
Db	895		
QY	961	MVKCWMIDSECRPRFRELVSFERSMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM	1020
Db	955		
QY	1021	GDLDVAEYLVPOQGFPCDPARGAGGVHHRHSSSTRSCGGDLTLGLEPSEEAAPRSP	1080
Db	1015		
QY	1081	LAPSEGAGSDVFDGDLGNAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTC	1140
Db	1075		
QY	1141	POPEYVNPQDVRPQPPSPREGPLPAARPAGATLERAKTLPSPKNGVVKDVFAGGAVENP	1200
Db	1135		
QY	1201	EYLTPOGGAAQPHPPPAFPAFNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVP	1260
Db	1195		
QY	1261	V 1261	
Db	1255	V 1255	

THIS PAGE BLANK (USPTO)

Nature 309, 270-273, 1984
A;Title: ATP-stimulated interaction between epidermal growth factor receptor and superoxide dismutase
A;Reference number: A38023; MUID:84191554; PMID:6325948
A;Contents: annotation; receptor activity
A;Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
R;Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.
Cell 59, 33-43, 1989
A;Title: Functional independence of the epidermal growth factor receptor from a domain of the EGF receptor
A;Reference number: A3331; MUID:90003233; PMID:2790960
A;Contents: annotation; internalization signal
C;Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor complex
C;Genetics:
A;Gene: GDB:EGFR
A;Cross-references: GDB:120610; OMIM:131550
A;Map position: 7p12.3-7p12.1
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphotyrosine
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-1210/Product: EGF receptor #status predicted <MAT>
F;25-645/Domain: extracellular #status predicted <EXT>
F;75-300/Domain: EGF receptor extracellular domain repeat <BE1>
F;390-600/Domain: EGF receptor extracellular domain repeat <BE2>
F;646-668/Domain: transmembrane #status predicted <TM>
F;669-1210/Domain: intracellular #status predicted <INT>
F;710-975/Domain: protein kinase homology <KIN>
F;718-726/Region: protein kinase ATP-binding motif
F;999-1046/Region: coated-pit mediated internalization signal
F;1047-1210/Region: inhibitory
F;128,175,352,414,444,528,603/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;745/Active site: Lys #status experimental

Query Match 46.0%; Score 3132; DB 1; Length 1210;
Best Local Similarity 49.1%; Pred. No. 9.6e-121;
Matches 626; Conservative 178; Mismatches 344; Indels 128; Gaps 22;

QY 11 LLLALLPPGAA-----STQVCTFNFTVFWLRVVKVSHLELYOGCOVQ 57
DB 14 LLAALCPASRALEBKVCQGTNSNKLTLQGTGFDFHFLSL-----QRMFNCEVVL 62
QY 58 GNLELTPLTNASLFQDIOQVGYVLIHNOVROVPLORLIRIVRGTQLFEDNYALAVL 117
DB 63 GNLEITVQRYNDLSFKLTQEVAGYVLIANTVERIPLENQLIIRGNMYENSALAVL 122
QY 118 DNGDPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILKWDIFKNN 177
DB 123 SNYD-----ANKTGLKELPMRNQLBILHGAVFNSNPALCNVESIQWRDIVSSDF 172
QY 178 QLALTLDITNRSRACHPCSPMKGSRGWGESSEDCQSLTRTVCAAGCA-RCKGPLPTDCC 236
DB 173 LSNMSMDFQNLHSGCQKDCSPNGSCWGAGEENCQKLTIIICAQCCSGRCRGSPPDCC 232
QY 237 HEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRVTFEGASCV 296
DB 233 HNQCAAGCTGPRSDCLVCKFRDEATCKTCTPPMLYNTPTTYQMDVNPBGKYSFGATCV 292
QY 297 TACPYNYSLTDVSGCTLVCLPHNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRV 356
DB 293 KKCPRNVVTDHGSVCRAAGADSSEM-EEDGVRKCKCEGPRCKVCGNGIGIGEFKDSLSI 351
QY 357 TSANIQEFACKKIFGSLAFIPSPFDGDPASNTAPLOPEQLQVPELTLEETGYLYISAMP 416
DB 352 NATNIKHFKNCTSISSGLDHLIPVAFRGDSFTHTPLDPLQBLDLTKVKEITGELLIOAWP 411
QY 417 DSLPLDSVFNLOVIRGIRLHNGAYSILTLQGLGISWGLRSLRGLSGSLALIHNTLCLF 476
DB 412 ENRTDLHAFENLEIIRGTQKHQGFSLAVSLNLSITLSGLSLKEISDGDVILSGNNKLCY 471
QY 477 VHTVPVQDLFRNPQALLHTANPEDRCVGEGLACHOLCARHGWCGPGTQCVCNCSQFLR 536
DB 472 ANTINMKKLGTSQKTKIISNRGENSKATGOVCHALCSPEGCWGPEDPCVSCRNVSR 531
QY 537 GQECVEBCRVLOGLPRRYVYVNRHCLPCHPECPQNGSVTCFGEADOCVCAHYKDPFPC 596
DB 537 GQECVEBCRVLOGLPRRYVYVNRHCLPCHPECPQNGSVTCFGEADOCVCAHYKDPFPC 596

DB 532 GRECDVKCKLLEGEPRFVENSCEIOCHPECLPQAMNITCTGRGPNCIOCAHYIDPHC 591
QY 597 VARCPGSKPDLGYMPTWKPDPDEGACQPCINCTHSCVDLDDKGCPAEQRASPLTISVS 656
DB 592 VKTCTPAGVGENNTL-VWKYADAGHVCHLCHPNCTYCTGCTPGLEGCTNGPKIP--SIAT 648
QY 657 AVVG-----LLVVVLGVVFGILIKRQOKIRKYTMRLLOTELVEPLTPSGAMPNOAMR 713
DB 649 GMVGAALLLVVALGIG--LFMRRRHIVKRLTLRLQLQRELVEPLTPSGEAPNQALLR 705
QY 714 ILKETELRVKVLGSGAGFCVYVYKIGVIGENVKIPVAIKVLRNTSPKANKELLDAYV 773
DB 706 ILKETEFKIKVLGSGAGFCVYVYKIGVIGENVKIPVAIKELREATSPKANKELLDAYV 765
QY 774 MAGVGSPPYVSRLLGICLTSTVQLVTLQMPYGCILLDHVRENRLGSGDGLNWCQIAKGM 833
DB 766 MASVDNPHVCRLLGICLTSTVQLITQMPFGCLLDYVREHKDNIGSQYLLNWCQIAKGM 825
QY 834 SYLEDVRLVHRDLAARNVLKSNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALES 893
DB 826 NYLEDRLVHRDLAARNVLKTPQHVKITDFGLAKLLGABEKEYHAEGGKVPKIKWMALES 885
QY 894 ILRRRTHGSDVMSYGVYVWELMTFGAKPYDGIIPABEIPDLLEKGERLPPOPICTIDVYM 953
DB 886 ILHRIYTHGSDVMSYGVYVWELMTFGSKPYDGIIPABEIPDLLEKGERLPPOPICTIDVYM 945
QY 954 IMVKCMIDSECRPRELVSERFARMARDPQRFVWIQ-NEDLGPASPLDSTFYRSLLEDD 1012
DB 946 IMVKCMIDADSPKRELIIEFSKWARDPQRLVILQGDRLHMLPSTDSNFRALMDEE 1005
QY 1013 DMGDLVDAEYLVPOQGFPCDPAPGAGGVVHRRSSSTRSGGDLTLGLEPSEEBAPR 1072
DB 1006 DMDDVDVDAEYLVPOQGF-----SSPSTSR 1031
QY 1073 SPLAPSEAGSDVFDGDLGMAAKGLQSLTHDPSPLORYSEDPVPLPSET--DGVPAP 1130
DB 1032 TPLLSLSLSTN--NSTVACIDRNGLOSCPEIKEDSFQRYSSDPTGALTEDSDDTFL-- 1087
QY 1131 LTCSPQEPYVQPDVPRQPPSPREGPLPAARPAATLERAKTLSPGKNGVVKVDFVAFGGA 1190
DB 1088 -----PVPEYINQ-SVPRFAGSVQNPVYHNOPLNP-----APSRDPHYQD--PHSTA 1132
QY 1191 VENPEYL-TPQGAAPQPHPPPAFSPADNLYYWDQ-----DP-----PERGAP 1233
DB 1133 VGNPEYLVNTVQ-----PTCVNSTFDSPAHWAQKSGHQSILDNPDYQDDFFPKAKP 1183
QY 1234 PSTFKGTPIHNPYLL 1249
DB 1184 NGIFKGS-TAENAEYL 1198

RESULT 5
A53183
epidermal growth factor receptor precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999
C;Accession: A53183; A43818; S24942; A28941; S45325; I49543
R;Luetteke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N.A.;
Genes Dev. 8, 399-413, 1994
A;Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor
A;Reference number: A53183; MUID:94170986; PMID:8125255
A;Accession: A53183
A;Molecule type: mRNA
A;Residues: 1-1210 <LUE>
A;Cross-references: GB:U03425
R;Avivi, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.
Oncogene 6, 673-676, 1991
A;Title: Comparison of EGF receptor sequences as a guide to study the ligand binding site
A;Reference number: A43818; MUID:91232866; PMID:2030916
A;Accession: A43818
A;Molecule type: mRNA
A;Residues: 1-714 <AVI>
A;Cross-references: GB:X59698

R;Lax, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennstr
Mol. Cell. Biol. 8, 1970-1978, 1988
A;Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou
A;Reference number: A27720; MUID:88261272; PMID:3260329
A;Accession: A27720
A;Molecule type: mRNA
A;Residues: 1-1223 <LAX>
A;Cross-references: GB:M20386
R;Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
Cell 41, 719-726, 1985
A;Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A;Reference number: A00643; MUID:95228222; PMID:2988784
A;Accession: A00643
A;Molecule type: mRNA
A;Residues: 585-1223 <NTL>
A;Cross-references: GB:M10066
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
specific protein kinase
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F;31-654/Domain: extracellular #status predicted <EXT>
F;81-307/Domain: EGF receptor extracellular domain repeat <EB1>
F;397-610/Domain: EGF receptor extracellular domain repeat <EB2>
F;655-677/Domain: transmembrane #status predicted <TM>
F;678-1223/Domain: intracellular #status predicted <INT>
F;719-984/Domain: protein kinase homology <KIN>
F;727-735/Region: protein kinase ATP-binding motif
F;136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #
F;192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F;687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;754/Active site: Lys #status predicted
F;1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 45.38; Score 3086.5; DB 1; Length 1223;
Best Local Similarity 48.48; Pred. No. 7e-119;
Matches 628; Conservative 175; Mismatches 349; Indels 145; Gaps 25;

QY 8 RWGLLALLPPGAA-----STOVCTENNFTVSPWLPVKVSAHSLLEYOGCQVQGNLE 61
DB 13 RGAIVLVLLGLVALCSAVEKKVCCQGTNNKLTLQGHVEDHFTSLQRMVNNCEVLNLE 72
QY 62 LTLPTNALSFLQDIQEVGYVLIHNVQVQLRQLRVGTQLPQEDNYALVLDNGD 121
DB 73 ITVEHNRDLTFLKTIQEVAGVYLIAMVDVIPLENLQIRGNVLYDNSFALVSNYH 132
QY 122 PLNNTTPTVTCASPGGLRELQRLSLEILKGVLIQRPQLCYQDTILWKDIFHNQNAL 181
DB 133 -MNKTO-----GLRELPMKRLSEILNGGVKISNNPKLCNMDTVLWNDIIDTSRK-PL 182
QY 182 TLID-TNRSRACHPCSPMKSGRCWGSSEDCSLTRTVCAAGCA-RCKGPLTDCHEQ 239
DB 183 TVLDFASNLSSCPKCHPNCNTEHCHWAGEQNCQLTKVICAQOCGRCRGKVPDCCNQH 242
QY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTPESMPNPEGRYTFGASCVTAC 299
DB 243 CAAGCTGPRESDCLACKFRDDATCKDTCPLVLNPTTYQMDVNPBGKYSFGATCVREC 302
QY 300 PNYLSTDVGSCTLVCPHLNQVETADGTORCSKSPCARVCYVGLGMEHLREVAVTSA 359
DB 303 PHNYVVTVDHSGSVRSCNVTDTYEV-EENGVRKCKCKDGLCSKVCNGIGIGLKGILSINAT 361
QY 360 NICEFAGCKKI FGLSLAPLPSFDGDPASNTAPLQEPLOVFEETLEETGVLYISANPDSL 419
DB 362 NISFKNCTKINGDVSLLPVAFLGDAFTPLDLPKKLDVFRIVKEISGFLLLQAWPDNA 421
QY 420 PDLVSFQNLQVIRGRILHNGAYSILTLQGLGISWLGRLSLRELGSGLALIHNTLHCFVHT 479
DB 422 TDLVAFENLEIIRGRTHQGOYSLAVNVNKLQISGLRLSEISDGDIAIMKNKVLVADT 481
QY 480 VPWDQLFRNPQALLHTANRPECEVGEGLACHQLCARGHCWGGPPTQCVCNCSQFLRGQE 539

RESULT 7

A47253

epidermal growth factor receptor, HER4 - human

C;Species: Homo sapiens (man)

C;Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999

C;Accession: A47253

R;Plozman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; N

Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993

A;Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal

A;Reference number: A47253; MUID:93189574; PMID:8383326

A;Accession: A47253

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: nucleic acid

A;Residues: 1-1308 <PLO>

A;Cross-references: GB:I07868; NID:g337359; PIDN:AAB59446.1; PID:g337360

A;Note: sequence extracted from NCBI backbone (NCBIP:126842)

C;Superfamily: epidermal growth factor receptor; protein kinase homology

C;Keywords: ATP; growth factor receptor
F;716-981/Domain: protein kinase homology <KIN>
F;724-732/Region: protein kinase ATP-binding motif

Query Match 43.6%; Score 2968.5; DB 2; Length 1308;
Best Local Similarity 45.2%; Pred. No. 4.8e-114;
Matches 612; Conservative 183; Mismatches 370; Indels 189; Gaps 30;

```
Qy 9 WGLLLALLPGAA-----STQVCTNNFTVSFWLVPKVSASHLE-----LYQGQV 56
Db 8 WVVSLVAAAGTVQPSQSQCAGTENKLS-----SLSDLEQQYRALRKYYENCEV 59
Qy 57 QGNLELYLPNNAISLFLODQVGVYLAHNOVROVPLQRLKRVGTOLFDENYALAV 116
Db 60 MGNLEITSIEHNRDLSFRVRETVYLVVALNFRYLPLENLRIIRGTIKLDYEDYALAI 119
Qy 117 LDNGDPLNNTPPVTVGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFPHN 176
Db 120 FLNRYKDGNF-----GLQELGLKLNTEILNGVGVVQNKFLCYADTIHQDVRNP 170
Qy 177 NQLALTILIDNRSRACHPCSPMKGRCWGSSEDCSLTRTVACAGC-ARCKGPLPTDC 235
Db 171 WPSNLTILVSTNGSSGCRCHKSTG-RCWGPTEHNCQTLTRTVCAEQDCRCYGPVYSDC 229
Qy 236 CHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTESMPNPEGRYTFGASC 295
Db 230 CHREACGCGSPKDTDFACNMFNDSGACVTQCQPTFYVPTTFQLEHNFNAKTYTGAF 289
Qy 296 VTACPYNLSTDVGSCTLVCPHLNHQVTAEDGTORCEKSKPCARVCYGLGMEHLREVR 355
Db 290 VKKCPHFV-VDSSSVCACRSPSSKMEV-EENGIRKMPCTDICKPCADGIGTGLMSAQT 347
Qy 356 VTSANIOEPAGCKIKFSLAFAPSPFDGDPASNTAPLOPQLOVFTLEBITGYLYISAW 415
Db 348 VDSNIDKFNCTKINGNLIFLVTGIHGDPPYNAIEADPEKLNVRFTVREITGFNLQSW 407
Qy 416 PDSLPDLVSFONLQVIRGRIHNGAYSITLQGLGSIWGLRSRLSGSLALIHNNHLC 475
Db 408 PPNMTDFSVFNLVTIGRVLYSGLSLILKQOQITSLQFOSLKEISAGNIYITDNLCL 467
Qy 476 FVHTVPMDQLFRPNHQALLTANRPEDECVGEGELACHQLCARGHCWGPPTQCVNCSQFL 535
Db 468 YHTINTTLFTSTINQIRIVIRDNKRAENCTAEGVMVCHLSCSSDCWGPDPDCLSCRRFS 527
Qy 536 RGQECVECRVLQGLPREYNARCLCHPECPQ-QNGSVTCFPEADQCVACHYKDP 594
Db 528 RGRICIESNLYDGEFEFENGSTCVCECPQCEKMEGDLTCHGPGPDNCTKSHFKDGP 587
Qy 595 FCVARCPGVKPDLSYMPIWKFPEEGACQPCPINCTHSCVDLDDKGC-----P 643
Db 588 NCVKCPDGLQGANF--IFKADPDRECHPCNCTQCGNGPTSHDCIYYPWTGHTLP 645
Qy 644 AEQASPLTISVAVV-GILLVVLVGVVFGILIKRROQKIRKTYMRLLQETELVEPLTP 702
Db 646 QHAR-TPL-IAAGVIGLFTLVGLTFAVYVRKSIK-KKRALRRFL-ETELVEPLTP 700
Qy 703 SCAMPNOAQRILKETELRKVKVLSGAFGVYIGWIPDGENKIPVAIKVLENTSPK 762
Db 701 SGTPAQOALRKELTKRKVKVLSGAFGVYIGWIPDGETVKIPVAIKILNETTGP 760
Qy 763 ANKEILDEAYVWAGVSPYVRLIGICTSTVQLVTOLMPYGCLLDHVRENRLGSDQL 822
Db 761 ANVEFMEALIMASDHPIVRLIGVCLSPILQVLTQMPHGCLLEVYHEKONIGSQL 820
Qy 823 LNCWQIAKMSYLEVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGG 882
Db 821 LNCVQIAKGMWYLEERLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGG 880
Qy 883 KVPILKWALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGRLP 942
Db 881 KMPILKWALECIHVRKETHQSDVMSYGVTVWELMTFGKPYDGIIPAREIPDLLEKGRLP 940
Qy 943 OPPICTIDVYIMVKCMWIDSECRPRELVSEFSRMARDPQRFVWIQNED-LGPASPLD 1001
```

```
Db 941 OPPICTIDVYIMVKCMWIDADSRPKFKELAAEFSSMARDPQRYLYIQGDRMKLPSND 1000
Qy 1002 STFYRSLLEDDMGDLVDAEYLVPOQGFCDPPAPGAGGVHRRSSSTRSGGDLTL 1061
Db 1001 SKFFQNLDEEDLEDMDAEYLVLP-QAFNIPPP-----IYTSRARIIDSNRS-----EI 1048
Qy 1062 GLEPSEEEAPRS-----PLAP-SEGAGSDVFDGDLGM 1092
Db 1049 GHSPPPAYTPMSONQFYRDGGFAAQGVSVYRAPRTSTIPEAPVAGATAEIPDDSCCN 1108
Qy 1093 GAAKGLQSLPTHPDPSPLQRYSEDTVPLPS-----ETDGYVAPLTCSPQPYVNPQDV 1145
Db 1109 GTLRKPVAVHQVQSDSTQRYSDPTVFAPERSPRGELDEEGYVTPMRDKPKQEVLNPE- 1167
Qy 1146 RPOPPSPRGPLPAAPAGATLERAKTLPSPKNGVVKVFAFGAVENPBYLTPQGGAA 1205
Db 1168 -----ENPFVSR-----KNGDLQ-----ALDNPEYHNASG----- 1194
Qy 1206 QPHPPPA-----FSPAFDNLVWDODPPERG 1231
Db 1195 ---PPKAEDYVNEPLVLTNFTANTLGAEYLNILSMPEKAKAFDNDPDYWNHSLPPRS 1251
Qy 1232 A--PPSTFKGTP-----AENPEYL 1249
Db 1252 TLQHPDYLQEYSTKYFVKQNGRIRPVAENPEYL 1285
```

RESULT 8

S06142
protein-tyrosine kinase (EC 2.7.1.112) mrk-y precursor - southern platyfish
N;Alternate names: epidermal growth factor receptor homolog; kinase-related transformin
C;Species: Xiphophorus maculatus (southern platyfish)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
C;Accession: S06142; S13809
R;Mittbrodt, J.; Adam, D.; Malitschek, B.; Maeueller, W.; Raulf, F.; Telling, A.; Robert
Nature 341, 415-421, 1989
A;Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu lo 50
A;Reference number: S06142; MUID:90015140; PMID:2797166
A;Accession: S06142
A;Molecule type: DNA
A;Residues: 1-1166 <WIT>
A;Cross-references: EMBL:X16891; NID:G65290; PIDN:CAA34770.1; PID:G65291
R;Adam, D.; Maeueller, W.; Scharlt, M.
Oncogene 6, 73-80, 1991

A;Title: transcriptional activation of the melanoma inducing Xmrk oncogene in Xiphophor
A;Reference number: S13807; MUID:91125882; PMID:1846957
A;Accession: S13809

A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>
A;Cross-references: EMBL:X56319; NID:G65284; PIDN:CAA39763.1; PID:G65285
C;Genetics:
A;Gene: mrk

A;Map position: Y
A;Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein; tyr
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>
F;707-972/Domain: protein kinase homology <KIN>
F;715-723/Region: protein kinase ATP-binding motif

Query Match 39.5%; Score 2689; DB 1; Length 1166;
Best Local Similarity 46.1%; Pred. No. 1.1e-102;
Matches 562; Conservative 156; Mismatches 372; Indels 129; Gaps 25;

```
Qy 48 ELYGCGVQVQGNLELTVLPTNASLFLQDQVGVYLAHNOVROVPLQRLVRQTOL 107
Db 51 KMTSGCNVNLLENLEITTTQENQDLSFLQSQVGVVLIAMNEVSTIPLVNLRLRQNL 110
Qy 108 FEDNYALAVLDNGDPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTI 167
```

Db 111 YEGNFTLLVMSNYOK- NPSSP- -DYVQVGLKQLQSLNTEILSGGVKVSHPNLLCNVETI 167
Qy 168 LMKDIFPHKNNQALTLITDYNRSRACHPCSPMCKGSRGWGESSEDCQSLTRIVTCAGGC-AR 226
Db 168 NWWDIVDKTSNPTWNLIPHAFCQKQCDHCVNGSCWAPGPGHCQKFTKLLCAEQCNRR 227
Qy 227 CKGPLPTDCCHQCAAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMNPDE 286
Db 228 CRGPKPTDCCNEHCAGGCTGPRATDCLACRFNDGDKTCDTPPKPKYDIVSHQVDPNP 287
Qy 287 GRYTFGASCVTACPVYVLTSDVSGCTLVCPPLHNEVTAEDGTORCEKSPCARVCYGLG 346
Db 288 IKYTFGACVKECSNTVWTE- GACVRSAGMLEVD- ENKRSCKPCDGVCPKCDGIG 345
Qy 347 MEHLREVRVTSANIOBFAGCKIFGLAFLPESFDGDPASNTAPLOPEQLQVPTILEE 406
Db 346 IGSLSNTIAVNSTIRSFNSCKINGDIIILNRNFSFGDPHYKIGTMDPEHLNLTITVKEI 405
Qy 407 TGYLYISAWPDSLPLSVFQNLQVIRGRILHNGAYS- LTLQGLIGISWGLRSLRELSG 465
Db 406 TGYLVIMWPNMTSLSVFQNLQVIRGRILHNGAYS- LTLQGLIGISWGLRSLRELSG 465
Qy 466 ALIHNTHLFCVHTVPMDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGP 525
Db 466 VILKNTLQLRVANTINRRULFRSDQSIYDART- -ENOTCNNECEDGCW- PGP 517
Qy 526 TQVNCQFRLRGQCEBCEVRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCV 585
Db 518 TMCVSCLVHVRGFCVASCNLLQGEPREA QV DGRVCVOHQECLVOTDSLTCYGPANCS 577
Qy 586 ACHYKDPFPCVACPCSGVKKDLSYMPIWKPDEBEGACQCPINCTHSCVDLDDKGPAP 645
Db 578 KSAHFODGPOQICPRCPHGLIGDGTLL- IWKYADRMGQCQCHQNTQCGSGPLSGCRGD 636
Qy 646 QRASPLTSIVSAYGILLVVLGVVGLILKRRQOKIRKYTMRELLQETLVEPLTSGA 705
Db 637 -IVSHSLAVGLVSLTITVALLIVLLRRRIK- RKTIRCLLQEKELVEPLTSGQ 694
Qy 706 MPNQAMRIKTELRKVKVLGSAFGTVYKGIPIPGENVKIPVAIKVLRNTPSRANK 765
Db 695 APQAFRLIKETEFKDRVLGSAFGTVYKGLWNPGENIRIPVAIKVLRNTPSRANK 754
Qy 766 EILDEAVMAGVSGPYVRLIGICLTSTVQLVTOLMPYGCLLDHVRNRRGLSGQDLLN 825
Db 755 EVLDEAVMASVDHPVRCRLLIGICLTSAVQLVTOLMPYGCLLDVYRQHERICQWLLN 814
Qy 826 CMQIAKMSYLEDLVRLDARNLVAKSNHVKITDFGLARLLDITDEYHADGKVP 885
Db 815 CVQIAKMNYLEERHLVRLDARNLVAKSNHVKITDFGLARLLDITDEYHADGKVP 874
Qy 886 IKWMALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPOLLEKGERLP 945
Db 875 IKWMALESILQWYTHQSDVMSYGVTVWELMTFGSKPYDGIPIAREIPOLLEKGERLP 934
Qy 946 ICTIDVIMVKMIDSECRPRELVSFBSRMDPQRVFIQNEBGLPASPLDSTFY 1005
Db 935 ICTIEVMIILKCMIDPSSRPRELVSFBSRMDPQRVFIQNEBGLPASPLDSTFY 991
Qy 1006 RSLLEDMDGLVDAAEYLPQOQFFCPDPAPAGAGVHHRRSSSTRSGGDLTLGLEP 1065
Db 992 SRLLEDMDGLVDAAEYLPQOQFFCPDPAPAGAGVHHRRSSSTRSGGDLTLGLEP 1019
Qy 1066 SEEAAPRSLAPSEAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPV- PLPSET 1124
Db 1020 - - - - -EPCIPPTGH- - - - -PVRENSITLRNISDPTQNALEKDL 1052
Qy 1125 DGYVAPLTCGQPYVNPQDVRPQ- - - - -PSPRE- - - - -GPLP- AARPAGATLBR 1171
Db 1053 DGH- - - - -EYVNPQSETSSRLSDIYNPNVEDLTDGWPVPSLSQEAETNFSRPE 1102
Qy 1172 TLPCKNGVVKVFAFGAVENPEYLTPOGGAAPQHPFPFAFNLYWDDPPERG 1231
Db 1103 YLNTNQNSL- - - - -PLVSSGSMDDPDY- - - - -YQAAF- - - - -LPQTG 1137

Qy 1232 APPESTFKGTPTAENPEYLG 1250
Db 1138 ALTGNGMFLPAENLEVLG 1156

RESULT 9

A36223

C:Species: Homo sapiens (man)
C:Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C:Accession: A36223; I59164
R:Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A:Title: Isolation and characterization of ERB3, a third member of the ERBB/epidermal
A:Reference number: A36223; MUID:90083234; PMID:2687875

A:Accession: A36223

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1342 <KRA>

A:CROSS-references: GB:M29366

R:Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, G.

Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990

A:Title: Molecular cloning and expression of another epidermal growth factor receptor-r

A:Reference number: I59164; MUID:90311312; PMID:2164210

A:Accession: I59164

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>

A:CROSS-references: GB:M34309; NID:G183390; PIDN:AAA35979.1; PID:G306841

C:Genetics:

A:Gene: GDB:ERBB3; HER3

A:Map position: 12q13-12q13

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homol

C:Keywords: ATP; phosphotransferase

F:707-972/Domain: protein kinase homology <KIN>

F:715-723/Region: protein kinase ATP-binding motif

Query Match 35.13%; Score 2408; DB 2; Length 1342;

Best Local Similarity 40.3%; Pred. No. 3.6e-91;

Matches 530; Conservative 191; Mismatches 447; Indels 146; Gaps 33;

Qy 5 ALCRWGLLLALLPPGAASTQVCTFNFTSFVLRVPKVSASHLELYGCGQVQGNLELY 64

Db 27 AVCP-GLNGLSVTGAENQYQTL- - - - -YKLYERCEVVMGNLEIVL 67

Qy 65 LPTNASLSFLQDIOEVQGVYVLIHNOVROPVLPORLRIVRGTQLFEDNYALVLDNGDPLN 124

Db 68 TGNADLSFLQWIREVTGYVLMANBFSTLPLNLRVVRGTQVYDGKFAIFVM- - - - -LN 122

Qy 125 NTPVTGASPGGLRELQRLSLTEILKGGVLIQNPOLCYQDITLWKDIFHKNQLALTLI 184

Db 123 YNT- - - - -NSSHAURQLRLTQLTEILSGGVYIEKNDKLCHMDTIDWRD- - - - -AEIV 175

Qy 185 DTRNRACHPCSPMCKGSRGWGESSEDCQSLTRIVTCAGGC- ARCKGPLPTDCCHQCAAG 243

Db 176 VKDNGRSCFPCHEVCKG- RCWGPGESEDCQTLTKTIAPQCNHGFPGFNPQCCHDECAGG 234

Qy 244 CTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMNPGRYTFGASCVTACPYNY 303

Db 235 CSGPQDTCFACRFNDGACVPRCPQPLVYNKLTQLQEPNPHTKYQYGCVCVASCPSHF 294

Qy 304 LSTDVSGCTLVCPPLHNEVTAEDGTORCEKSPCARVCYGLMEHLREVRVTSANIOE 363

Db 295 V-VDQTSVCRACPPDKMEVD- - - - -KNGLKMCPCGGLCPKACEGTGSG- - - - -SRFQTVDSNIDG 350

Qy 364 FAGCKIFGSLAPLPSFDGDPASNTAPLOPEQLQVPTILEETGYLYISAWPDSLPLDS 423

Db 351 FVNCTKILGNLDLITGLNGDPWHKIPALDPEKLNVRFTVREITGYLNTQSWPHHNF 410

Qy 424 VFQNLQVIRGRILHNGAYS- LTLQGLIGISWGLRSLRELSGGLALIHNTHLFCVHTVP 482

Db 411 VFSNLTITGSRSLYNGFSLIMKLNLTSLGPRSLKEISAGRIYISANRQLCYHHSLNW 470
Qy 483 DQLFRNPHQALLHTA-NRPEDECVGEGLACHOLCARGHCWPGTQCWNSQFLRGQECV 541
Db 471 TKVLRGTEERLDIKHNRPRDCVAEGKVCPLCGSSGCGWPGGQCLSCRNYSRGVGV 530
Qy 542 EECRVQLQPREYVNAHCLPCHPECOPOGNSVTCFGEADQCVACAHYKDPFPCVACRP 601
Db 531 THCNFLNGEPREFAHAEACSCHEPCQMEGTATCNGSGSDTCAQCAHFRDGPCHVSCP 590
Qy 602 SGVKPDLSPYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCAPBQRA-----SPLTSIVSA 657
Db 591 HGVLG--AKGPIYKPDVQNECRCHENCCTGCKGPELQDCLGQTLVLIGKTHLTMTLV 648
Qy 658 VVGLLVVVLGVVFGIILKRQOKIR-KYTHRRLLQETELVEPLTPSGAMPNQOMRLK 716
Db 649 IAG--LVVIFMMLGGTFLYWRGRIONKRAMRRYLERGESIEPLDPS-EKANKVLARIFK 705
Qy 717 ETELKRVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAG 776
Db 706 ETELKRVKVLGSGVFGTVHKGWVLPAGESIKIPVICIKVEDKSGRQSFQAVTDHMLAIGS 765
Qy 777 VGSYVSRLLGICHTSTVQLTQMPYXGCLLDHVRNRLGSLQDILLNMCQIAKMSYL 836
Db 766 LDHAHIVRLGLCPGSSQLVTVPLGLSLLDHVRQHRGALGPQLLLNMGVQIAKGMVYL 825
Qy 837 EDVLRVHRLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNMALESILR 896
Db 826 BEHGMVHRLAARNVLLKSPSQVQVADFGVADLLPPDDKQLLYSEAKTPIKNMALESIFH 885
Qy 897 RRFTHQSDVMSYGVTVNEMTFGAKPYDGI PAREIPDLLEKGERLPPOPICTIDVYMW 956
Db 886 KXYTHQSDVMSYGVTVNEMTFGAEYAGLRLAEVDPDLLEKGERLAQOICTIDVYMWV 945
Qy 957 KCMWIDSECRPRFRELSEFSEFARMARDPQRFVVIQNEIDLGPA---SPLDSTFYRSLLEDD 1013
Db 946 KCMWIDENIRPTEKLANEFTRMARDPPRYLVIKRES-GPGIAPGPEPHGLTNKKLEVE 1004
Qy 1014 MGLVDABEYLVPOQFCPPDPAPGAGGVHHRHSSTRSGGDLTLIGLEP-SEEAAPR 1072
Db 1005 LEPELDLDLDEABED-----NLATTGLSALSPLVGTNLNRRGSG 1045
Qy 1073 SPLAPSEGAGSDVFDGLMGAAKGLSLPTH--PSPLQRYSEDTVPPLP-----SETD 1125
Db 1046 SLLSPSGY-MPMNQNLGESCQSAVSGSSERCPRVSLH-----PMRGCLASESE 1098
Qy 1126 GYVA-----PLTCSQPE-----YVNPQDVRPQPPSPREG- 1156
Db 1099 GHVTGSEAEQLQEKVSMCRSRSRSPRPGDSAYHSQRHSLLTPTVPLSPGLSEEDVNG 1158
Qy 1157 --LPAARPAGATLERAKTLSP-KGNVY-----KDVFAFGAVENPEYLTPOGGAPOP 1207
Db 1159 YVMPDTHLKGTPSSREGTLSSVGLSLGTDEED-----EYEVNRRRRHSP-P 1209
Qy 1208 HPPPAFSPFDNLVYWD-----QPPPERGAPPSTFKTPTAENPEYL 1249
Db 1210 HPPRPSLEELGYEYMDVGSLSASLSGTSQCPLHPVIMPATGTTDEDEYEM 1263

RESULT 10
JC4387

epidermal growth factor receptor homolog precursor - rat

N:Alternate names: ErbB3 protein; HER3 protein

C:Species: Rattus norvegicus (Norway rat)

C>Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998

C:Accession: JC4387

R:Heilwer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.

Gene 165, 279-284, 1995

A>Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.

A:Reference number: JC4387; MUID:96096535; PMID:8522190

A:Accession: JC4387

A:Molecule type: mRNA

A:Residues: 1-1339 <HEL>

A;Cross-references: GB:U29339; NID:g915389; PID:g915390

A;Experimental source: liver

A;Note: The authors translated the codon AAC for residue 369 as Thr and GTT for residue 370. This protein is a functional heregulin receptor that transduces signals to the

C;Genetics:

A;Gene: ErbB3

C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homol.

C;Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-1339/Product: epidermal growth factor homolog #status predicted <MAT>

F;640-659/Domain: transmembrane #status predicted <TM>

F;705-970/Domain: protein kinase homology <KIN>

F;713-721/Region: protein kinase ATP-binding motif

F;939,1051,1156,1194,1196,1219,1257,1273,1286,1325/Binding site: phosphate (Tyr)

Query Match

34.2%; Score 2327.5; DB 2; Length 1339;

Best Local Similarity 40.6%; Pred. No. 6.9e-88;

Matches 519; Conservative 170; Mismatches 422; Indels 167; Gaps 34;

Qy 5 ALCRWGLLLALLPPGAASQVCTFNNFTVSFMLRVPKVSASHLEYGCGCOVQGNLELY 64
Db 27 AVCP-GTLNGLSVTGADNQYQL-----YKLEYKEVVMGNLEIVL 67
Qy 65 LPTNASLSFLQDIQEVQGVYLIAHNOVRQVPLQRLRIVRGTLQFEDNYALAVLDNGDPLN 124
Db 68 TGHNADLSFLQIREVTAYVIVAMNEFSLPLENLAVRGTVQYDGFKAIFVM----LN 122
Qy 125 NTPVTGASPGURELQRLSLTEILKGVLIQRNPOLCYQDTILKWDIFHNNQLAULTI 184
Db 123 YNT-----NSHALRQLKFTQLTEILSGGVYIERNDKLCMDTIDWRDVR---GAEIV 175
Qy 185 DTNRSRACHPCSPMKSGSRGSESSDCSLTQTVTCAGGC-ARCKGPLPTDCHEOCAAG 243
Db 176 VRNNGANCPCPEHCVKG-RCWGPDDCQLTKTTCAPQCNCRGFGPNPQCCHDECAGG 234
Qy 244 CTGPKHSDDLCLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNY 303
Db 235 CSGQDTDFACRRFNDGACVPCPEPLVYNKLTQLQEPNHTKYQYGGCVASCPCPNF 294
Qy 304 LSTVSGSCTLVCPHINQEVTAEDGTORCEKSPCARVCYGLQMEHLREVRATVSANIQE 363
Db 295 V-VQDTFCVRACPPDKMEVD-KHGLMKCEPCGGLCPKACEGTGSG--SRYQTVDSSNIDG 350
Qy 364 FACKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEETITGVLIYISAWPDSLPLDS 423
Db 351 FVNCTKLGLMDFLITGLNVDPWHKIPALDPEKLVNFRVREITGYLINIOSWPHMNF 410
Qy 424 VFQNLQVIRGIRLHNGAYS-LTLQGLGISWGLSLRSLRSLGSLALIHNNTHLCFVHTVPW 482
Db 411 VFSNLTITGSRSLYNGFSLIMKLNLTSLGPRSLKEISAGRIYISANRQLCYHHSLNW 470
Qy 483 DQLFRNPHQALLHTA-NRPEDECVGEGLACHOLCARGHCWPGTQCWNSQFLRGQECV 541
Db 471 TKVLRGTEERLDIKHNRPRDCVAEGKVCPLCGSSGCGWPGGQCLSCRNYSRGVGV 530
Qy 542 EECRVQLQPREYVNAHCLPCHPECOPOGNSVTCFGEADQCVACAHYKDPFPCVACRP 601
Db 531 THCNFLNGEPREFAHAEACSCHEPCQMEGTATCNGSGSDTCAQCAHFRDGPCHVSCP 590
Qy 602 SGVKPDLSPYMPIWKFPDEEGACQPCINCTHSC---VDLDDKGCFAEQRASPLTSIVSAV 659
Db 591 HGVLG--AKGPIYKPDVQNECRCHENCCTGCKGPELQDCLGQTLVLIGKTHLTMTLV 648
Qy 660 GILLVVVLGVVFGIILKRQOKIR-KYTHRRLLQETELVEPLTPSGAMPNQOMRLKET 716
Db 649 IAG--LVVIFMMLGGTFLYWRGRIONKRAMRRYLERGESIEPLDPS-EKANKVLARIFK 705
Qy 719 ELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGV 778
Db 706 ELRKVLKVLGSGVFGTVHKGWVLPAGESIKIPVICIKVEDKSGRQSFQAVTDHMLAIGS 765
Qy 779 SPYVSRLLGICHTSTVQLTQMPYXGCLLDHVRNRLGSLQDILLNMCQIAKMSYLED 838
Db 706 ELRKVLKVLGSGVFGTVHKGWVLPAGESIKIPVICIKVEDKSGRQSFQAVTDHMLAIGS 765
Qy 779 SPYVSRLLGICHTSTVQLTQMPYXGCLLDHVRNRLGSLQDILLNMCQIAKMSYLED 838

Db 715 QYTAIGY-----CAASPPRSSKITANLD::: : : : :
Qy 669 VVFGI-LIKRQOKIRYKT--MRLLOTELVEPLTPSGAMPNOAMRILKETELRKVKV 725
Db 756 ILCVVYICQKQAKKETVMTWALSREDSEPLRPSNIGANLCKLRIVKDAELRGKV 815
Qy 726 LGSAGFTVYKGIWIPGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVSGPYVSR 785
Db 816 LGMAGFRVYKGVWVPGENVKIPVAIKELLKSTGABSSEFLREAYIMASEEHVNLKL 875
Qy 786 LGICLTSTVOLATOLMPYVGLDHRNCRGLSGODLLNMCWIAKMSYLEDLRLVHRD 845
Db 876 LAVCMSSQMLITOLMPYVGLDHRNCRKIGSKALLNMQSTAKMSYLEDLRLVHRD 935
Qy 846 LAARNVLVK---SPNHVKITDFGLARLDIDETEHADGGKVPKMKWMALESILRRRTHQ 902
Db 936 LAARNVLRLLAGEDH---DFGLAKLLSSDSNEYKAAGKMPKWLALCALCIRNREVTSK 991
Qy 903 SDVMSYGVVWELMTFGAKPYDGIIPAREIPDLLEKGBRLPOPPICITIDVYMWKMWID 962
Db 992 SDVMAFGTIIWELLTFQORPHENIPAKDIIPDLIEVGLKLEQPEICSLDIYCTLLSCWHL 1051
Qy 963 SECPRERELVSERSMARDPQRFVIONEDLG--PASPLDSTFYRSILLEDD---DMGDL 1017
Db 1052 AAMRPTFKQLTTVFAEFARDPGRYLAITLGDKFTRLPA-----YTSQDEKDLIRKLAPT 1104
Qy 1018 VDAEYLVPQGGFCPPAPGAGGMVHRRSSSTRSGGDLTUGLEPSEEEAP----- 1071
Db 1105 TDGSEATKDDYLOPKAALGPS-----HRTDCT-----DEMPKLNRYC 1143
Qy 1072 RSLAPSEAGSDVFDG---DLGMAAGLQSLTHDPSLQRYSEDPTVPLPSETDGYV 1128
Db 1144 KPSNKNSSGDDERDSAREVGVNLR-----LDLPVDEDDYL 1182
Qy 1129 APLTCSQPEYVQPDVPPPPREGPLPAARPAGATLERAKTLSPGKGVVKDVPFG 1188
Db 1183 MP-TCQPGPNNNMN-----NPNQNNMAAVGAAGYM-----DLIGVP 1220
Qy 1189 GAVENPEYL---TPQGGAAPOPH-----PPAFSP-AFDNLYYWD 1224
Db 1221 VSDNPEYLLNAQTLGVGSEPIPTQITIGIPWGGPGTMEVKVMPGSEPTSSDHEYYND 1279

RESULT 14

S35745
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
C;Species: avian erythroblastosis virus
C;Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
C;Accession: S35745
submitted to the EMBL Data Library, March 1993
A;Reference number: S35743
A;Accession: S35745
A;Molecule type: DNA
A;Residues: 1-544 <VEN>
A;Cross-references: EMBL:X12707
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
F;135-400/Domain: protein kinase homology <KIN>
F;143-151/Region: protein kinase ATP-binding motif
F;170/Active site: Lys #status predicted

Query Match 24.2%; Score 1647; DB 2; Length 544;
Best Local Similarity 54.9%; Pred. No. 1.6e-60;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

Qy 578 GPEADQCVACAHYKDPFPCVACPSGVKPDLSYMPIWKFDEGACQPCINCTHSCVDL 637
Db 1 GP--DHCMKCAHFIDGPHCVKACPAVLGENDTL-VMKYADANAVCOLCHPNCNTRGCKGP 57

Qy 638 DDKGCFAEGRASPLTSIVSAV--GILLVVVLGVVFGILIKRROOKIRKYTMRLRLQETEL 696
Db 58 GLEGCP---NGSKTPSIAAGWGLLCLVVGVLGIGLYLRRR-HIVKRTLRLLQEREL 113
Qy 697 VEPITSGAMPNOAMRILKETELRKVKVLSGAFCTVYKGIWIPGENVKIPVAIKVL 756
Db 114 VEPITSGEAPNOAHRIKETELRKVKVLSGAFCTVYKGIWIPGENVKIPVAIKVL 173
Qy 757 ENTSPKANKEILDEAYVMAGVSGPYVSRLLGICLTSTVOLVQTLMPYVGLDHRNCR 816
Db 174 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVOLITOLMPYVGLDHRNCR 233
Qy 817 LGSQDILLNMCWIAKMSYLEDLRLVHRDLAARNVLKSNHVKITDFGLARLDIDETE 876
Db 234 IGSQYLLNMCVQIAKGMVLEERHVMVHRDLAARNVLKTPQHVKITDFGLAKQLGADEKE 293
Qy 877 YHAGGKVPKMKWMALESILRRRTHQSDVMSYGVVWELMTFGAKPYDGIIPAREIPDLLE 936
Db 294 YHAGGKVPKMKWMALESILHRIYTHQSDVMSYGVVWELMTFGSKPYDGIIPAREIPDLLE 353
Qy 937 KGBRLPOPPICITIDVYMWKMWIDSECRPFRELVSERSMARDPQRFVIO--NEDLG 995
Db 354 KGBRLPOPPICITIDVYMWKMWSDADSPKRELIAEFSKWARDPPRVLYIQGDRMH 413
Qy 996 PASPLDSTFYRSILLEDDMGDLVDAEYLVPQGGFCPPAPGAGGMVHRRSSSTRSG 1055
Db 414 LPSPTDSKF/RTLMBEEDMEDI VDAEYLVPHQGF-----NSPST--- 454
Qy 1056 GGDLTGLSESEEARSPPL-----APSEGAGSDVFDGDLGMAAGLQSLTHDPSPLQ 1110
Db 455 -----SRTPLLSLSATSNNSATNCIDRNGG-----H----- 481
Qy 1111 RYSEDPVPLPSETDGYVAPLTCSPQPEYVQPDVPPPPREGPLPAARPAGAT--LER 1169
Db 482 -----PVREDGFL-----PAPEYVQ--LMPKPESTAMVQNIYVLSLTISK 523
Qy 1170 AKTLSPGKGVVKDVPFAFGGAVENPEYL 1197
Db 524 LPIDRSYQN-----SHSTAVDNPEYL 544

RESULT 15

S00727
kinase-related transforming protein (erbB) (EC 2.7.1.1.-) - avian erythroblastosis virus
C;Species: avian erythroblastosis virus
C;Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
C;Accession: S00727
R;Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
Oncogene Res. 1, 265-278, 1987
A;Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mutan
A;Reference number: S00727; MUID:88217326; PMID:2897102
A;Accession: S00727
A;Molecule type: DNA
A;Residues: 1-545 <SCO>
A;Cross-references: EMBL:X06943
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; phosphotransferase
F;135-400/Domain: protein kinase homology <KIN>
F;143-151/Region: protein kinase ATP-binding motif

Query Match 24.1%; Score 1640; DB 2; Length 545;
Best Local Similarity 54.9%; Pred. No. 3.1e-60;
Matches 345; Conservative 69; Mismatches 122; Indels 92; Gaps 15;

Qy 578 GPEADQCVACAHYKDPFPCVACPSGVKPDLSYMPIWKFDEGACQPCINCTHSCVDL 637
Db 1 GP--DHCMKCAHFIDGPHCVKACPAVLGENDTL-VMKYADANAVCOLCHPNCNTRGCKGP 57
Qy 638 DDKGCFAEGRASPLTSIVSAV--GILLVVVLGVVFGILIKRROOKIRKYTMRLRLQETEL 696
Db 58 GLEGCP---NGSKTPSIAAGWGLLCLVVGVLGIGLYLRRR-HIVKRTLRLLQEREL 113

```
Qy 697 VEPLTPSGAMPNOQMRTLKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLR 756
Db 114 VEPLTPSGEAPNOAHLRLKETEFKKVKVLGFGAGFTVYKGLWIPGEKVTIPVAIKELR 173
Qy 757 ENTSPKANKEILD EAYVMAGVSPVSRLLGICLTSTVQLVTQLMPYGCCLLDHVRENRR 816
Db 174 EATSPKANKEILD EAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCCLLDYIREHKDN 233
Qy 817 LGSODLLNWCQIAKGSYLEEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 876
Db 234 IGSQYLLNWCQIAKGMNYLEERHLVHRDLAARNVLKTPQDVKITDFGLAKQLGADEKE 293
Qy 877 YHADGGKVPISKMALESTLRRRFTHQSDVWSYGVTVWELMTFGAKPYDGPAREIPDLLE 936
Db 294 YHAEKKVPISKMALESTLHRIYTHQSDVWSYGVTVWELMTFGSKPYDGPASEISSVLE 353
Qy 937 KGERLPQPPICTIDVYIMVKCWMIDSECRPRPRELVSEFSRMARDPQRFVVIQ-NEDLG 995
Db 354 KGERLPQPPICTIDVYIMVKCWMSDADSRPKFRELIAEFKSMARDPPRYLVIQDERMH 413
Qy 996 PASPLDSTFYRSILLEDDMGDLVDAEYLVPOQGFCEPDPAPGAGGMVHHRRSSSTRSG 1055
Db 414 LPSPTDSKFYRTLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
Qy 1056 GGDLTGLPSEEEAPRSP-----APSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPQLQ 1110
Db 455 -----SRTPLLSLSATSNNSATNCIDRNG-----H----- 481
Qy 1111 RYSEDPTVPLPSETDGVYVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAGAT-LER 1169
Db 482 -----PVREDGFL-----PAPEYVQ--LMPKKPSTAMVQNIYNYISLTAISK 523
Qy 1170 AKTLPKGKNGVVKDVAFPGGAVENPEYL 1197
Db 524 LPMDSRYN-----SHSTAVDNPEYL 544
```

Search completed: July 22, 2003, 09:07:53
Job time : 30.9062 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2304 Seconds
(without alignments)
5088.033 Million cell updates/sec

Title: SEQ4-5-25-14

Perfect score: 6812

Sequence: 1 MELAALCRWGLLALLPPGA.....TFKGTPTAENPEYLGDPVP 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6694	98.3	1255	1 ERB2 HUMAN	P04626 homo sapien
2	5882	86.3	1257	1 ERB2_RAT	P06494 rattus norv
3	5879.5	86.3	1254	1 ERB2_MESAU	Q60553 mesocricetu
4	3130	45.9	1210	1 EGFR HUMAN	P00533 homo sapien
5	3112	45.7	1210	1 EGFR MOUSE	Q01279 mus musculus
6	2968.5	43.6	1308	1 ERB4 HUMAN	Q15303 homo sapien
7	2949	43.3	1308	1 ERB4_RAT	Q62956 rattus norv
8	2705.5	39.7	1167	1 XMRK_XIPWA	P13388 xiphophorus
9	2417	35.5	1342	1 ERB3 HUMAN	P21860 homo sapien
10	2350.5	34.5	1339	1 ERB3_RAT	Q62799 rattus norv
11	1931	28.3	1426	1 EGFR DROME	P04412 drosophila
12	1749.5	25.7	634	1 ERBB_ALV	P00534 avian leuko
13	1703	25.0	604	1 ERBB_AVIER	P00535 avian eryth
14	1630	23.9	540	1 ERBB_AVIEU	P11273 avian eryth
15	1587	23.3	703	1 EGFR_CHICK	P13387 gallus gall
16	1297	19.0	1323	1 LT23 CAEL	P24348 caenorhabdi
17	1142.5	16.8	245	1 ERB2 MOUSE	P70424 mus musculus
18	724	10.6	1363	1 ILPR_BRALA	O02466 branchiosto
19	700	10.3	1382	1 INSR_HUMAN	P06213 homo sapien
20	699.5	10.3	1607	1 MIPF_LYMST	Q95410 lymphocyst
21	695	10.2	1300	1 IRR_MOUSE	Q9wt14 mus musculus
22	694.5	10.2	1372	1 INSR_MOUSE	P15208 mus musculus
23	694	10.2	1383	1 INSR_RAT	P15127 rattus norv
24	694	10.2	1477	1 HTK7_HYDAT	Q25197 hydra atten
25	687	10.1	1297	1 IRR_HUMAN	P14616 homo sapien
26	684.5	10.0	1300	1 IRR_CAVPO	P14616 cavia porce
27	645	9.5	1367	1 IGIR_HUMAN	P08069 homo sapien
28	635	9.3	1373	1 IGIR_MOUSE	Q60751 mus musculus
29	631.5	9.3	1370	1 IGIR_RAT	P24062 rattus norv
30	625.5	9.2	1390	1 INSR_AEDAE	Q93105 aedes aegypt
31	621.5	9.1	2146	1 INSR_DROME	P09208 drosophila
32	615	9.0	987	1 EPB4_HUMAN	Q54760 homo sapien
33	591.5	8.7	984	1 EPB1_CHICK	Q07494 gallus gall

RESULT 1

ID	ERB2_HUMAN	STANDARD;	PRT;	1255 AA.
AC	P04626;			
DT	13-AUG-1987 (Rel. 05, Created)			
DT	13-AUG-1987 (Rel. 05, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)			
DE	(p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell			
DE	surface receptor HER2) (MLN 19).			
GN	ERBB2 OR HER2 OR N3L OR NEU.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OX	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
NCBI_TaxID=9606;				
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86118663; PubMed=3003577;			
RA	Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,			
RA	Saito T., Toyoshima K.;			
RT	"Similarity of protein encoded by the human c-erbB-2 gene to			
RT	epidermal growth factor receptor."			
RL	Nature 319:230-234 (1986).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86070181; PubMed=2999974;			
RA	Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,			
RA	McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,			
RA	Franccke U., Levinson A., Ullrich A.;			
RT	"Tyrosine kinase receptor with extensive homology to EGF receptor			
RT	shares chromosomal location with neu oncogene."			
RL	Science 230:1132-1139 (1985).			
RN	[3]			
RP	SEQUENCE OF 737-1031 FROM N.A.			
RX	MEDLINE=86016729; PubMed=2995967;			
RA	Semba K., Kanata N., Toyoshima K., Yamamoto T.;			
RT	"A v-erbB-related protooncogene, c-erbB-2, is distinct from the			
RT	c-erbB-1/epidermal growth factor-receptor gene and is amplified in a			
RT	human salivary gland adenocarcinoma."			
RL	Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501 (1985).			
RN	[4]			
RP	VARIANTS VAL-654 AND VAL-655.			
RX	MEDLINE=93194196; PubMed=8095488;			
RA	Ehsani A., Low J., Wallace R.B., Wu A.M.;			
RT	"Characterization of a new allele of the human ERBB2 gene by allele-			
RT	specific competition hybridization."			
RL	Genomics 15:426-429 (1993).			
CC	-I- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,			
CC	ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A			
CC	POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-			
CC	ALPHA AND AMPHIREGULIN.			
CC	-I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein			
CC	tyrosine phosphate.			
CC	-I- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS			
CC	(POTENTIAL).			
CC	-I- SUBCELLULAR LOCATION: Type I membrane protein.			

ALIGNMENTS

34	589.5	8.7	977	1	EPB4_MOUSE	Q03145 mus musculus
35	588	8.6	987	1	EPB4_MOUSE	P54761 mus musculus
36	588	8.6	1114	1	RET_HUMAN	P07949 homo sapien
37	584.5	8.6	976	1	EPB2_HUMAN	P29317 homo sapien
38	583.5	8.6	984	1	EPB1_RAT	P09759 rattus norv
39	579	8.5	902	1	EPBB_XENLA	Q91736 xenopus lae
40	577.5	8.5	984	1	EPB1_HUMAN	P54762 homo sapien
41	576.5	8.5	985	1	EPBA_XENLA	Q91571 xenopus lae
42	573.5	8.4	1053	1	PAK1_CHICK	Q00944 gallus gall
43	569	8.4	1068	1	PAK1_XENLA	Q91738 xenopus lae
44	567	8.3	757	1	HT16_HYDAT	P53356 hydra atten
45	563	8.3	1052	1	PAK1_MOUSE	P34152 mus musculus

QY	717	ETELRKVKV	LGSGAGFTYKGIWITPDGENVKIPVAIKVLRENTSPKANKETLDEAYVWAG	776
Db	717	ETELRKVKV	LGSGAGFTYKGIWITPDGENVKIPVAIKVLRENTSPKANKETLDEAYVWAG	776
QY	777	VGSPVVSRL	LIGICLTSTVQLVTQLMPPYGCLLDHHVRENRRGLSGODLLNWCQAKGMSYL	836
Db	777	VGSPVVSRL	LIGICLTSTVQLVTQLMPPYGCLLDHHVRENRRGLSGODLLNWCQAKGMSYL	836
QY	837	EDVELVHRD	LAARNVLKSPNHVKITDGLARLLDIDETEYHADGGKVPKMWALESIILR	896
Db	837	EDVELVHRD	LAARNVLKSPNHVKITDGLARLLDIDETEYHADGGKVPKMWALESIILR	896
QY	897	RRFTHQSDVMS	GYVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMV	956
Db	897	RRFTHQSDVMS	GYVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMV	956
QY	957	KCMWIDSECR	PRFRELVSERWARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDMDG	1016
Db	957	KCMWIDSECR	PRFRELVSERWARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDMDG	1016
QY	1017	LVDAAEYLV	PQQGFFCDDPAPGAGMVHRRHSSTRSGGGDLTLGLPSEEEAPRSPLA	1076
Db	1017	LVDAAEYLV	PQQGFFCDDPAPGAGMVHRRHSSTRSGGGDLTLGLPSEEEAPRSPLA	1076
QY	1077	PSEGAGSDV	FDGDLGMGAAGKLSLTHDPSPLQRYSEDPTVPLPSTDGYVAPLTCSPQ	1136
Db	1077	PSEGAGSDV	FDGDLGMGAAGKLSLTHDPSPLQRYSEDPTVPLPSTDGYVAPLTCSPQ	1136
QY	1137	PEYUNQDVR	PPSPREGPLPAARPAAGATILERAKTILSPKNGVVKDVFAGGAVENPEY	1196
Db	1137	PEYUNQDVR	PPSPREGPLPAARPAAGATILERAKTILSPKNGVVKDVFAGGAVENPEY	1196
QY	1197	LTPGGGAAP	QPHPPAPFSPAFDNLVYWDQPPERGAPSTFKGTPTAENPEYLGLDVVP	1255
Db	1197	LTPGGGAAP	QPHPPAPFSPAFDNLVYWDQPPERGAPSTFKGTPTAENPEYLGLDVVP	1255
RESULT 2				
ID	ERB2	RAT	STANDARD;	PRT; 1257 AA.
AC	P06494;			
DT	01-JAN-1988	(Rel. 06, Created)		
DT	15-DEC-1998	(Rel. 37, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		
DE	Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)			
DE	(p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor receptor-related protein).			
GN	ERBB2	OR NEU.		
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
NCBI	TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Neuroblastoma;			
RX	MEDLINE=86118662; PubMed=3945311;			
RA	Bargmann C.I., Hung M.-C., Weinberg R.A.;			
RT	"The neu oncogene encodes an epidermal growth factor receptor-related protein.";			
RL	Nature 319:226-230(1986).			
RN	[2]			
RP	SEQUENCE OF 852-905 FROM N.A.			
RC	TISSUE=Sciatic nerve;			
RX	MEDLINE=91222560; PubMed=2025425;			
RA	Lai C., Lemke G.;			
RT	"An extended family of protein-tyrosine kinase genes differentially expressed in the vertebrate nervous system.";			
RL	Neuron 6:691-704(1991).			
RN	[3]			
RP	STRUCTURE BY NMR OF 650-668.			
RX	MEDLINE=92195181; PubMed=1346763;			
RA	Gullick W.J., Bottomley A.C., Lofts F.J., Doak D.G., Mulvey D.;			

```
FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT MOD_RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 86.3%; Score 5882; DB 1; Length 1257;
Best Local Similarity 86.3%; Pred. No. 3.6e-306;
Matches 1088; Conservative 54; Mismatches 109; Indels 10; Gaps 4;

Qy 1 MELALCRWGLLLALLPPGAASSTOVCTFNFTVSFWLRVVKVSASHLE----LYQGCOVV 56
Db 1 MELAAWCRWGLLLALLPPGAGTAGTCGTDMK----LRLPASPETHLDMLRHLVYQGCQVV 56

Qy 57 QGNLELTYPNALSLSFLQDIQEVQVYLAHNOVROVPLQRLRIVRGTOIFEDNYALAV 116
Db 57 QGNLELTYPNALSLSFLQDIQEVQVYLAHNOVRRVPLQRLRIVRGTOIFEDNYALAV 116

Qy 117 LONGDPLNNTTPTVT-GASPGGLRELQRLSLTEILKGGVLIQORNPLQCYQDTILMKDIFHK 175
Db 117 LDRNDPQDNVAASTPGRTPEGLRELQRLSLTEILKGGVLIQORNPLQCYQDMWLWKDVERK 176

Qy 176 NNQALTLIDNRSRACHPCSPMKGRCRGWESSEDQSQSLTRTVACGCGARCKGPLPTDC 235
Db 177 NNQALPVDIIDNRSRACPPCAPACKNHCWGESPEQCQILTGITCTSGCARCKGRLEPTDC 236

Qy 236 CHEOCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTPESMPNPEGRVTFGCASC 295
Db 237 CHEOCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTPESMHNPEGRVTFGCASC 296

Qy 296 VTACPNYLSLDVSGSCTLVCPHNLHNOVTAEDGTQRCCKSPCARVCYGLGMEHLREVR 355
Db 297 VTTCPNYLSLEVSGSCTLVCPNNQVTAEDGTQRCCKSPCARVCYGLGMEHLRGARA 356

Qy 356 VTSANIQEFAGCKIFGSLAPLPSFDGDPASNTAPLOPEQLOVFLEBITGYLYISAW 415
Db 357 ITSDNVQEFDCCKIFGSLAPLPSFDGDPSSGIAPLRPEQLOVFLEBITGYLYISAW 416

Qy 416 PDSLPDLVSFONLOVIRGRILHNGAYSILTQGLGISWGLRLSRLRELGSGLALIHNNTHLC 475
Db 417 PDSLRDLVSFONLRIIRGRILHDGAYSILTQGLGIHSLRLSRLRELGSGLALIHNNHLC 476

Qy 476 FVHTVPWDQLFRNPHQALLHTANPEPE-CVGEGLACHOLCARGHCWGPQPTQVCNCSQF 534
Db 477 FVHTVPWDQLFRNPHQALLHSGNRPEEDLCVSSGLVCNLSLCAHCHGWGPQPTQVCNCSHF 536

Qy 535 LRGECEVEECRVLOGLPREYNARHCLPCHPECPQNGSVTCFGEADQVACAHYKDDP 594
Db 537 LRGECEVEECRVKGLPREYVSDRXCLPCHPECPQNSSETCFGEADQVACAHYKDDSS 596

Qy 595 FCVARCPGSKVPKDLVSYMPIWPKFPEEGACQPCPNTCHSCVDLDDKGCPCAEORASPLTSI 654
Db 597 SCVARCPGSKVPKDLVSYMPIWPKFPEEGICQPCPNTCHSCVDLDERGCPAEORASPTFI 656

Qy 655 VSAVVGILLVVLGVWFGLIKRRQKIRKYTMERLLQETELVEPLTPSGAMPNQAMRI 714
Db 657 IATVVGILLVFLVWVVGILLIKRRQKIRKYTMERLLQETELVEPLTPSGAMPNQAMRI 716

Qy 715 LKETELRKVKVLGSGAGTGVYKGIWDGENVKIPVAIKVLRNTPSKANKEILDEAYVM 774
Db 717 LKETELRKVKVLGSGAGTGVYKGIWDGENVKIPVAIKVLRNTPSKANKEILDEAYVM 776

Qy 775 AGVGSPPVSRLLGICLTSTVQLVTPYGLLDHVRNRRGLSGDQLLNWCMQIAKGMS 834
```

```
Db 777 AGVGSPPVSRLLGICLTSTVQLVTPYGLLDHVRNRRGLSGDQLLNWCMQIAKGMS 836
Qy 835 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESI 894
Db 837 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESI 896
Qy 895 LARRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYMI 954
Db 897 LARRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYMI 956
Qy 955 MVKCMWIDSECRPFRELVSERFMRARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1014
Db 957 MVKCMWIDSECRPFRELVSERFMRARDPQRFVVIQNEIDLGPSSPMDSTFYRSLLEDDDM 1016
Qy 1015 GDILVDAEYLVPQGGFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLLEPSEEEAPRSP 1074
Db 1017 GDILVDAEYLVPQGGFFCPDPPTGTGTARRHRRSSSTRSGGDLTLGLLEPSEEGPRSP 1076
Qy 1075 LAPSEGAGSDVFDGLGMAAGLQSLPTHDPSPLOQYSEDPTVPLPSETDGVVAPLTCS 1134
Db 1077 LAPSEGAGSDVFDGLGMAAGLQSLPTHDPSPLOQYSEDPTVPLPSETDGVVAPLTCS 1136
Qy 1135 POPEYVQNDVRRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENP 1194
Db 1137 POPEYVQNDVRRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENP 1196
Qy 1195 EYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPRGAPSTPKGTPTAENPEYLGLDVP 1254
Db 1197 EYLVPRREGTASPPHPPAFSPAFDNLVYWDQDPPRGAPSTPKGTPTAENPEYLGLDVP 1256
Qy 1255 V 1255
Db 1257 V 1257

RESULT 3
ERB2_MESAU
ID_ERB2_MESAU STANDARD; PRT; 1254 AA.
AC Q60553;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
DE ERB2 OR NEU.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RS TISSUE=Nerve;
RC MEDLINE=94193007; PubMed=7908275;
RA Nakamura T., Uehijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.,
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255(1994).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```



```
FT DISULFID 620 628 BY SIMILARITY.
FT MOD_RES 624 636 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT MOD_RES 680 680 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE)
FT CARBOHYD 128 128 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 19 19 C -> S (IN REF. 2).
FT CONFLICT 539 539 C -> W (IN REF. 5).
FT CONFLICT 991 991 L -> F (IN REF. 4).
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).
SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46DF2D2F5 CRC64;

Query Match 45.78; Score 3112; DB 1; Length 1210;
Best Local Similarity 49.18; Pred. No. 1.4e-158;
Matches 630; Conservative 169; Mismatches 352; Indels 132; Gaps 24;

QY 11 LLLALLPPGAA-----STQVCTFNFTVFWLRVPRKVSASHLELYGCGVQV 57
DB 14 LLLTALCAAGALBEKKVCOGTSNLTQTGFEDHFLSL-----QRMVNCVEVL 62
QY 58 GNLELYLPTNASLFLQDIOEVQVYLIHVNQVQVPLQRLIRVGTQFLFDNYALAVL 117
DB 63 GNLEITYVQRYNDLSPLKTOEVAGVYLIANTVERIPLENLQIIRGNALYENTYALAIL 122
QY 118 DNGDPLNTPVTGASPGGLREQLRLSLTEILKGGVLIQRNPOLYQDTILWKDI----F 173
DB 123 SN-----YGTNRTGRLFMRLNQLBILGAVRFSNNPILCNMDTTQWRDVIQNVF 172
QY 174 HKNQLALTLIDNRSRACHPCSPMKSGSRCSWESSEDCOSLRTVTCAGCA-RCKGPLP 232
DB 173 MSNWSMDL-----QSHSSCKPKDPCSPGSCWGGEENCOKLTKIICAOQCCHRCGRSP 228
QY 233 TDCHEQCAAGCTGPKHSDCLACILHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFG 292
DB 229 SDCHNCAAGCTGPRSDCLVCQKQFDEATCKDTCPLMLXNPTTVQMDVNEPGKVSFG 288
QY 293 ASCVTACPNYSLTDVSGSCTLVCLPHNQEVTAEDGTQRCBCKSPCARVCYGLGMEHLRE 352
DB 289 ATCVKCPRYNVVTDHSGSCVRACGPDYEV-EEDGIRKCKKCDGPKRCKVNCGIGIEFKD 347
QY 353 VRAVTSANIOEAGCKKIFGSLAFLPESPDGPNASNTAPLOPQLOVFTLEITGLVLYI 412
DB 348 TLSINATNIKFKYCTAISGLHLVPAFGDSFTPTPDPRELETLTKVKEITGFLLI 407
QY 413 SAWPDSLDLSVFQNLQVIRGRHNGAYSILTLQGLGISWGLRSLRSLRELGSGLALIHNT 472
DB 408 QAMPDNWTDLHAFENLEIRGTQKQGFQSLAVVGLNITSGLRSLKEISDGDVILSGNR 467
QY 473 HLCFVHTVPDWQJLFRNPHQALLHTANRPEDECVEGEGLAGHOLCARGHCWGPGTQCVCNS 532
DB 468 NLCYANTINWKLFGTPTNQTIKMNRNRAEKDCAVNHVNCNPLCSSESGCWPPEPRDCVSCQ 527
QY 533 QFLRGQCEVECVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKD 592
DB 528 NVSRGECVBKMLIEGEPREFENSEICQCHPECPQAMNITCTGRGPNQNCIQCAHYD 587
QY 593 PPFVCARCPGVKPDLSYMPIDEGACQPCPCINCTHSCVDLDDKCGPQASPLT 652
DB 588 GPHCVKTCPAGINGENNTL-VMKYADANNVCHLCHANCTYGCAGPGQCGCEVWPSGPKIP 646
QY 653 SIYSAVVGIILLVVVLGVVFGI-LIKRQQKIRKYTWRRLLQETELVEPLTPSGAMPNQAQ 711
```

```
DB 647 STATGVGGLLFTIW-VALGIGLFMRRRHVRKRTLRLLQERLEVEPLTPSGEAPNQA 705
QY 712 MRILKETELRKVKVLGSGAGFTYKGIWIPDGNWVPIPAVKVLRENTSPKANKEILDEA 771
DB 706 LRILKETEFKIKVLGSGAGFTYKGIWIPDGNWVPIPAVKELRATSPKANKEILDEA 765
QY 772 YNAGVGSVPVSRLLGICLTSTVOLVQTMPLYPGCLLDHVRENRGRIGSQDQLLNCMQIAK 831
DB 766 YMASVDNEFVCRLLGICLTSTVOLITQTMPLYPGCLLDVYREHKDNIGSQVLLNWCQIAK 825
QY 832 GNSYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEYHADGGKVPICKMAL 891
DB 826 GMYLEDRRLVHRDLAARNVLYKTPQHVKITDFGLAKLLGAEKEYHAEGKVPICKMAL 885
QY 892 ESILRRFRFHQSQVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDV 951
DB 886 ESILHRIYTHQSQVMSYGVTVWELMTFGSKPYDGIIPASDISSILEKGERLPQPPICTIDV 945
QY 952 YIMVVKCWMIDSECRPRFRELSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLE 1010
DB 946 YIMVVKCWMIDADSRPKFRELILEFSKMAKDPOQVLYIQGDERMHLPSPTDSNFYRALMD 1005
QY 1011 DDDMGDLVDAEYLVPOQFFCPDPAPGAGMVHHRSSSTRSGGDLTLGLEPSEEEA 1070
DB 1006 EEDMEDVDADADEYLIPOQGGF-----NSPST----- 1031
QY 1071 PRSPLAPSGAGSDVDPDGLGMAAKGLOSPLTHDPSLQRYSEDDTVPPLPSET--DGYV 1128
DB 1032 SRTPLLSLSLATS-----NSTVACINRNGSCRVKEDAFQRYSSDPTGAVTENIDDAFL 1087
QY 1129 APLTCSQPQYVNVQDVRPQPPSPREGPLPAARPAAGATLAKTLSPGKNGVVKDVFAFG 1188
DB 1088 -----PVPEYVNO-SVPRKRPAGSVQNPVYHQPLHP-----APGRDLHYON--PHS 1130
QY 1189 GAVENPEYL-TQCGGAAPQHPHPPAPSPAFDNLVYWDQ-----DP-----PERG 1231
DB 1131 NAVGNPEYLNTAQ-----PTCLSSGFNSPALMIQKSHQMSLDNPDYQODFFPKET 1181
QY 1232 APPSTFKGFTAEENPEYLGLDVP 1254
DB 1182 KPNGIFKG-FTAENAEYLRVAPP 1203

RESULT 6
ERB4 HUMAN STANDARD; PRT; 1308 AA.
ID ERB4 HUMAN
AC Q15303;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
DE (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).
DE ERB4 OR HER4.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP SEQUENCE FROM N.A. (ISOFORM JM-A).
RC TISSUE=Breast carcinoma;
RX MEDLINE=93189574; PubMed=8383326;
RA Plowman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,
RA Foy L., Neubauer M.G., Shoyab M.;
RT "Ligand-specific activation of HER4/p180erbB4, a fourth member of the
RT epidermal growth factor receptor family.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM JM-A AND JM-B).
RC TISSUE=Fetal brain;
RX MEDLINE=97476287; PubMed=9334263;
RA Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plowman G.D.,
RA Klagsbrun M.;
```

"A novel juxtamembrane domain isoform of HER4/ErB4. Isoform-specific tissue distribution and differential processing in response to phorbol ester." J. Biol. Chem. 272:26761-26768 (1997).

CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND NTA. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.

CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN.

CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.

CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS (POTENTIAL).

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; JM-A (SHOWN HERE) AND JM-B; ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.

CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART, KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM, PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS, LUNG, SALIVARY GLAND, AND PANCREAS.

CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES.

CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; L07868; AAB59446.1; -.

DR HSP; P11362; 1FGK.

DR Genew; HGNC:3432; ERBB4.

DR MIM; 600543; -.

DR InterPro; IPR000494; EGFR_L domain.

DR InterPro; IPR000719; Euk_pkinase.

DR InterPro; IPR002174; Furin-like.

DR InterPro; IPR001245; Tyr_pkinase.

DR InterPro; IPR004019; YLP motif.

DR Pfam; PF00069; pkinase; 1.

DR Pfam; PF00757; Furin-like; 1.

DR Pfam; PF01030; Recep_L domain; 2.

DR Pfam; PF02757; YLP; 2.

DR ProDom; PD000001; Euk_pkinase; 1.

DR SMART; SMO0261; FU; 4.

DR SMART; SMO0219; TyKc; 1.

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.

DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal; Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation; Alternative splicing.

FT SIGNAL 1 25

FT CHAIN 26 1308

FT DOMAIN 26 651

FT TRANSMEM 652 675

FT DOMAIN 676 1308

FT DOMAIN 186 334

FT DOMAIN 496 633

FT DOMAIN 718 985

FT NP_BIND 724 732

FT BINDING 751 751

FT ACT_SITE 843 843

FT DISULFID 189 197

FT DISULFID 193 205

FT DISULFID 213 221

FT DISULFID 217 229

FT DISULFID 230 238

RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.

EXTRACELLULAR (POTENTIAL).

POTENTIAL.

CYTOSOLASMIC (POTENTIAL).

CYS-RICH.

CYS-RICH.

PROTEIN KINASE.

ATP (BY SIMILARITY).

ATP (BY SIMILARITY).

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

FT DISULFID 234 246 BY SIMILARITY.

FT DISULFID 249 258 BY SIMILARITY.

FT DISULFID 262 289 BY SIMILARITY.

FT DISULFID 293 304 BY SIMILARITY.

FT DISULFID 308 323 BY SIMILARITY.

FT DISULFID 326 330 BY SIMILARITY.

FT DISULFID 503 512 BY SIMILARITY.

FT DISULFID 507 520 BY SIMILARITY.

FT DISULFID 523 532 BY SIMILARITY.

FT DISULFID 536 552 BY SIMILARITY.

FT DISULFID 555 569 BY SIMILARITY.

FT DISULFID 559 577 BY SIMILARITY.

FT DISULFID 580 589 BY SIMILARITY.

FT DISULFID 593 614 BY SIMILARITY.

FT DISULFID 617 625 BY SIMILARITY.

FT DISULFID 621 633 BY SIMILARITY.

FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

FT CARBOHYD 138 138 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 174 174 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 181 181 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 253 253 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 358 358 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 410 410 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 473 473 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 495 495 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 548 548 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 576 576 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 620 620 N-LINKED (GLCNAC...) (POTENTIAL).

FT VARSPLIC 626 648 NGPTSHDCIVYPTMGHSTLTPQHA -> IGSSIEDICGLMD (IN ISOFORM JM-B).

FT SQ SEQUENCE 1308 AA; 146807 MW; 5E4AE80985D88761 CRC64;

Query Match 43.6%; Score 2968.5; DB 1; Length 1308;

Best Local Similarity 45.2%; Pred. No. 6.7e-151;

Matches 612; Conservative 183; Mismatches 370; Indels 189; Gaps 30;

QY 9 WGLLLALLPPGAA-----STQVCTFNNFTVSFWLRVFKVSASHLE-----LYQCGQV 56

DB 8 WVVVSLVAAGTVQPSDSQSVACATENKLS-----SLSDLEQQYALRKTYENCEV 59

QY 57 QGNLELTYPNTASLSFLQDIQEVQGVLIHQAQVQVPLQRLRIVRGTPQEDNVALAV 116

DB 60 MGNLEITSHEHNRDLSFLRSRVETGYVLVALNQFRLPLENRIIRGKLYEDRALAI 119

QY 117 LONGDPLNNTTPVTGASPGGLRELQLRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFKN 176

DB 120 FLNRYKDNF-----GLQELGLKNTLILGGVYVDQNKFLCYADVTHHQDIVRNP 170

QY 177 NQALTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSLTRTVCAAGC-ARCKGPIPTDC 235

DB 171 WPSNLTALVSTNGSSGCGRCHKSCGTG-RCWGPTENHCQTLTRTVCAEQCDGRCYGVYSDC 229

QY 236 CHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGYTFGASC 295

DB 230 CHREAGGSGPKDTCFACMFNDSGACVTCQCPQTFVFNPTTFQLEHFNAYTYGAF 289

QY 296 VTACPNYLTSDVGSCTLVCPHLNQEVTAEDGTCRCEKCKPCARVCYGLGMBHLREVRA 355

DB 290 VKKCPHFV-VDSSSCVCRACPPSKMEV-EENGFKMCKPCTDICPKACDGICTGSLMSAQ 347

QY 356 VTSANTQEFAGCKKIFGSLAFLEPESFDGPASNTAPLOPEQLOVFTLEBITGYLISAW 415

DB 348 VDSNIDKFINCTKINGNLIFLVTHGIDPYNIAEIDPEKLVFRTVREITGFLNIQSW 407

QY 416 PDSLPLSVFQNLQVIRGRILHNGAVSLTLOQGISWLGRLSRELGLALHHTHLC 475

DB 408 PNMWTFVSFNLTITGGRVLYSGLSLLILKQGITSLQFQSLKEISAGNIYTDNSNLC 467

QY 476 FVHTVPMQDLFRNPHQALLHTANRPEDECVGEGACHQLCARGHCGPGTQCVCNCSQFL 535

```

Db 468 YYHTINTLSTINQIRIVIRDRNKAENCTAEGMVCNHLCSDDGCGGPPQDCLSCRRFS 527
Qy 536 RGQCEVCECRVQLGPLEYVNRHCLPCHPQCP-QNGSVTCFGEADQCCACAHYKDP 594
Db 528 RGRICIESCNLYDGEFREFENGSIQVECDPQCEKMEGDLTCHGPGDNCCTKCSHFQGP 587
Qy 595 FCVACPSGVKPDLSYMPIKFPDEGACQCPINCTHSCVDLDDKC-----P 643
Db 588 NCVEKCPDGLQGANF--IFKYADPDRECHFCNCTQGCNGPSTHCDIYYPWTGHSTLP 645
Qy 644 AEQASPLTSIVSAV--GILLVVLGVVFGILIKRROOKIRKYTWRLLOETELVEPLTP 702
Db 646 QHAR-TPL--IAAGVIGLFLIVIGLTFVAVYVRKSIK-KKRALRFL-ETELVEPLTP 700
Qy 703 SGAMPNQAQRILKETELRKVKVLGSGAFGVYKGIWIPDGENKIPVAIKVLRNTPK 762
Db 701 SGTAPNQAQRILKETELRKVKVLGSGAFGVYKGIWIPDGENKIPVAIKVLRNTPK 760
Qy 763 ANKEILDEAYMAGVSGPYSTRILGICLTSTVQLVTLMPYGCLLDHVRENRLGSQL 822
Db 761 ANVEFDEALIMASMDPHLVRLLGVCLSPITQILVTQLMPHGCCLLYVHEHKONIGSQL 820
Qy 823 LNCWQQLAGMSYLEVRLVHRDLAARNVLKSNHVKITDFGLARLLDIDETEVHADGG 882
Db 821 LNCVQIAGMVLLEERLVRDLAARNVLKSNHVKITDFGLARLLDIDETEVHADGG 880
Qy 883 KVPKMWALSIILRRRTHOSDVMSYGVYVWELMTFGAKPYDGIPIAREIPDLLEKGRLP 942
Db 881 KMPKMWALSIILRRRTHOSDVMSYGVYVWELMTFGAKPYDGIPIAREIPDLLEKGRLP 940
Qy 943 OPPCTTIDVYMWKMWIDSECRPRELVSFSRMARDPQRFVWIONED-LQASPLD 1001
Db 941 OPPCTTIDVYMWKMWIDSECRPRELVSFSRMARDPQRFVWIONED-LQASPLD 1000
Qy 1002 STFYRSLEDDMDGLVDAEYLVPOQGFPCPDPAAGAGGVHRRHSSSTRSGGDLTL 1061
Db 1001 SKFQNLDEEDLEDMDAEYLV-PQAFNIPP-----IYTSRAIDNRS-----EI 1048
Qy 1062 GLEPSEEAPRS-----PLAP-SGAGSDVFDGDLGM 1092
Db 1049 GHSPPPAYTMSGNQFVYRDGFAAQGVSVYRAPTSITPEAPVAQATAEFDSCCN 1108
Qy 1093 GAAKGLQSLTHDPSLQVSEDTVPLPS-----ETGYVAPLTCSPQEPVYVPODV 1145
Db 1109 GTLRKVPAPHVQEDSSTQRYASDTPVAPERSPRGELDEGYTPMDKPKQYLNVE- 1167
Qy 1146 RPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVAFGAVENPEVLTPOGGAAP 1205
Db 1168 -----ENPFVSR-----KNGDLQ-----ALDNPEYHNASNG--- 1194
Qy 1206 QPHPPA-----FSPADNLYWQDPPERG 1231
Db 1195 ---PPKADEYVNEPLNTFANTGLKAEYLVKNNILSMPEKAKAFDNPYWNHSLPPRS 1251
Qy 1232 A--PPSTFKGTP-----AENPEYL 1249
Db 1252 TLQHPDYLOEYSTKYFYKONGRIRPIVAENPEYL 1285

RESULT 7
ERB4_RAT
ID ERB4_RAT STANDARD; PRT; 1308 AA.
AC Q62956; Q922N7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN ERB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
RA Marchionni M.A., Kelly R.A.;
RT "Neuregulins promote survival and growth of cardiac myocytes.
RT persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes.";
RL J. Biol. Chem. 273:10261-10269(1998).
RN [2]
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).
RN [3]
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659(1997).
CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTA. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF041838; AAC08899.1; -.
CC EMBL; U25231; AAC53051.1; -.
CC HSSP; P11362; 1FGK.
CC InterPro; IPR000494; EGFR_L_domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC InterPro; IPR004019; YLP_motif.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF01030; Recep_L_domain; 2.
CC Pfam; PF02757; YLP; 2.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 4.
CC SMART; SM00219; TyrKC; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
RN SIGNAL 1 25 POTENTIAL.

```


FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 652 675 POTENTIAL.
FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 496 633 PROTEIN KINASE.
FT DOMAIN 718 985 ATP (BY SIMILARITY).
FT NP BIND 724 732 ATP (BY SIMILARITY).
FT BINDING 751 751 BY SIMILARITY.
FT ACT_SITE 843 843 BY SIMILARITY.
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 246 BY SIMILARITY.
FT DISULFID 249 258 BY SIMILARITY.
FT DISULFID 262 289 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.
FT DISULFID 308 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 559 577 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 614 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 621 633 BY SIMILARITY.
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 138 138 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC.) (POTENTIAL).
FT CONFLICT 1062 1062 S -> N (IN REF. 3).
FT CONFLICT 1080 1082 PYT -> SYR (IN REF. 3).
SQ SEQUENCE 1308 AA; 146957 MW; D944BB0996A08B41 CRC64;

Query Match 43.3%; Score 2949; DB 1; Length 1308;
Best Local Similarity 45.0%; Pred. No. 7.3e-150;
Matches 610; Conservative 190; Mismatches 379; Indels 176; Gaps 30;

Qy 1 MELA-ALCRWGLLL--ALLPPGAASQVCTFNFTVFWLVRPVKVSASHLE-----L 49
Db 1 MKLATGLWVWSSLLVAARTVQPSASQSVCACTENKLS-----SLSDLEQQYRALRKY 52
Qy 50 YQGCVQVGNLELYLPFNASLSFLQIQVQGVLYIAHNVQVQVPLQRLRIVRGTLQFPE 109
Db 53 YENCEVWGNLEITSIEHNRDLSLSIREVTGVLYVALNQFRLPLENRIIRGTLYE 112
Qy 110 DNYALAVLDNGDPLNNTPVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILW 169
Db 113 DRYALALFLNVRKGNP-----GLQELGLKNLLEILNGGVYVQNKFLCYADIHWH 163
Qy 170 KDIFHKNQLALTIDNRSRACHPCSPMKGSRGWGSSSDCQSLTRTVCAGGC-ARCK 228
Db 164 QDIVRNWPNWNTLVSTIGSGGRCCHKSCGTG-RCWGPTEHNCQTLTRTVCAGCDGRCY 222
Qy 229 GPLPTDCHEOCAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGR 288
Db 223 GPVYSDCCHRECAGCGSGPKDTCDFACMNFNDSGACVTCQCPQTFVYNPNTTFQLEHFNNAK 282

Qy 289 YTFGASVCTACPYNYLSTDVGSCTCLPLHNOQVTAEDGTQRCCKSKPCARVCYGLGME 348
Db 283 YTYGAFCKVKCPHFV-VDSSSCVRACPPSSKMEV-EENGIKMCKPCTDICKACDGIITG 340
Qy 349 HLEVRATVSANTQEFAGCKKIPGSLAFIPESPDGPASNTAPLOPELOQVFTLEBITG 408
Db 341 SLMSAQTVDSNIDKFINCTKINGNLIFLVGTIGHDPYNAIDAIDPEKLNVFTRBITG 400
Qy 409 YLYISAWPDSLPLSVFQNLQVIRGILHNGAVSLTQGLGISWGLRSILRELSGLALI 468
Db 401 FLNIQTWPNMTDFSVFSNLVTIGGRVLYSGLSLLILKQOGITSLQFSKESAGNIYI 460
Qy 469 HNHTHLCFVHTVPWDLFRNPQALHTANRPEDECVGEGSLACHQLCARGCHGWPGTQC 528
Db 461 TDNSNLCCYHTINWTTLSFVINOIRVIRONRAENCTAEGMVCNHLCSNDGCGWPGPDQC 520
Qy 529 VNCQFLRGQECVECRVLOGLPREVYNARHCLUPCHPEQOP-QNGSVTCFPGPADQCVAC 587
Db 521 LSCRRFSRGKICIESCNLYDGEFEPENGSIQCEVDSQCEKMEDEGLLTCHGPGDNCTKC 580
Qy 588 AHYKDPFPFCVARCPSGVKPDLSYMPIWKFPDEBEGACQPCPINCTHSCVDLDDKGC----- 642
Db 581 SHFKDGNPCVCKPCDVLQGANSE--IFKYADQDRECHPCHPNCTQCGNGTSHDCIYYPW 638
Qy 643 -----PAEQRASPLTSIVSAVV-GILLVVVLGVWFGLIKIRKQKIRKYTMRLLOETE 695
Db 639 TGHSTLTPQHAR-TPL--IAAGVIGGLFILVIMALTAVVYRRKSIIK-KKALARRFL-ETE 693
Qy 696 LVEPLTPSGAMPNQAQMRILKETELRKVKVGLSGAGFTGVYKGIWIDGENVKIPIVAIKVL 755
Db 694 LVEPLTPSGTAPNQAQRLILKETELRKVKVGLSGAGFTGVYKGIWVPEGETVKIPIVAIKIL 753
Qy 756 RENTSKANKEILLDEAVVMAGVGSPIVSRLLGICLTSTVOLVTLQMPYCGLLDHHVNRG 815
Db 754 NETTGPKNVFEFDEALIMASVDHPLVRLVLLGCLSPITQLVTLQMPHGCCLLYVHEHDK 813
Qy 816 RLGSQDLLNWCMIQAKGMSYLEDLVLRDLAARNVLKSPNHNKIDTDFGLARLLDIDET 875
Db 814 NIGSOLLNWCVOIAGMWMYLEERLVRDLAARNVLKSPNHNKIDTDFGLARLLGDEK 873
Qy 876 EYHADGGKVPKIMWALESILRRRFTHQSDVMSYGVTVWELMTGAKPYDGIPIAREIPDLL 935
Db 874 EYNADGGKMPKIMWALECIHYRKFTHQSDVMSYGVTVWELMTGAKPYDGIPIAREIPDLL 933
Qy 936 EKERLPQDPICITDVMIMVKCMWIDSECRFRELVSFMRARDPQRFVIONED-L 994
Db 934 EKERLPQDPICITDVMIMVKCMWIDSECRFRELVSFMRARDPQRFVIONED-L 993
Qy 995 GPASPLDSTFYRSLLEDDDDMDGLVDAEYLVPOQGFPCDP----- 1035
Db 994 KLPSPNDSKEFFONLLDEEDLEMDMDAEYLVLP-QAFNIPPIYTSRTRIDSNRSEIGHSP 1052
Qy 1036 ----APGAGMWHHRHSSTRSGGDLTLGLBPSBEEAPRSPPLASBEGAGSDVFDGLG 1091
Db 1053 PPATYTMGSGQFVYQGGFATQGG--MMPYATTTSTIPEAPVA--QGATAEMFDDSCC 1107
Qy 1092 MGAAGLQSLPHTDPSPLQRYSEDPTVPLPS-----ETDGYVAPLTCSPQBYVYNQPD 1144
Db 1108 NGTLRKPVPVPHVQEDSTQRYSDPTVFAPERNPRALDEEGYMTBMHDKPKQBYLNPVE 1167
Qy 1145 VRQPPSPRGPPLAARPAAGATLERAKTLSPGKNGVVKVDFAFGGAFAVENPEYLTPOGGAA 1204
Db 1168 -----ENPFVSR-----KNGDLQ-----ALDNEYHSASSG-- 1194
Qy 1205 PQPHPPA-----FSPAFDNLYYWDQDPPER 1230
Db 1195 ----PKADEYVNEPLYLNTFTNALGNAEYMKNSLLSVPEKAKAFDNPYWNHSLPPR 1250
Qy 1231 GA--PSTFKGT-----AENPEYL 1249
Db 1251 STLQHPDYLOEYSTKYFKYKONGRIRPIVAENPEYL 1285


```

RESULT 8
XMRK_XIPWA
ID XMRK_XIPWA STANDARD; PRT; 1167 AA.
AC P13388;
DT 01-JAN-1990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
GN XMRK OR TU.
OS Xiphophorus maculatus (Southern platyfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX NCBI_taxid=8083;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90015140; PubMed=2797166;
RA Wittbrodt J., Adam D., Malitschek B., Maueier W., Raulf F.,
RA Telling A., Robertson S.M., Scharltl M.;
RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
RT inducing Tu locus in Xiphophorus.";
RL Nature 341:415-421(1989).
RN [2]
RP REVISION TO 515.
RA Scharltl M.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X16891; CAA34770.2; --
CC PIR: S06142; S06142.
CC HSSP: P11362; IFGK.
CC InterPro: IPR000494; EGFR_L_domain.
CC InterPro: IPR000719; Euk_Pkinase.
CC InterPro: IPR002174; Furin-like.
CC InterPro: IPR002290; Ser_thr_pkinase.
CC InterPro: IPR001245; Tyr_pkinase.
CC Pfam: PF00069; pkinase; 1.
CC Pfam: PF00757; Furin-like; 1.
CC PRINTS: PR01030; Recep_L_domain; 2.
CC ProDom: PD000001; Euk_pkinase; 1.
CC SMART: SM00261; FU; 5.
CC SMART: SM00220; S_TKC; 1.
CC SMART: SM00219; TyrKc; 1.
CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
CC Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
CC SIGNAL
CC 1 25
CC CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE
CC KINASE.
CC DOMAIN 26 642 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 643 665 POTENTIAL.
CC DOMAIN 666 1167 CYTOPLASMIC (POTENTIAL).
CC DOMAIN 710 977 PROTEIN KINASE.
CC NP_BIND 716 724 ATP (BY SIMILARITY).
CC BINDING 743 743 ATP (BY SIMILARITY).
CC ACT_SITE 835 835 BY SIMILARITY.

```

```

FT DISULFID 195 204 BY SIMILARITY.
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 220 228 BY SIMILARITY.
FT DISULFID 224 236 BY SIMILARITY.
FT DISULFID 237 245 BY SIMILARITY.
FT DISULFID 241 253 BY SIMILARITY.
FT DISULFID 256 265 BY SIMILARITY.
FT DISULFID 269 296 BY SIMILARITY.
FT DISULFID 300 311 BY SIMILARITY.
FT DISULFID 315 330 BY SIMILARITY.
FT DISULFID 333 337 BY SIMILARITY.
FT DISULFID 504 513 BY SIMILARITY.
FT DISULFID 508 521 BY SIMILARITY.
FT DISULFID 524 533 BY SIMILARITY.
FT DISULFID 537 553 BY SIMILARITY.
FT DISULFID 556 569 BY SIMILARITY.
FT DISULFID 560 577 BY SIMILARITY.
FT DISULFID 593 615 BY SIMILARITY.
FT DISULFID 618 626 BY SIMILARITY.
FT DISULFID 622 634 BY SIMILARITY.
FT CARBOHYD 114 114 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 144 144 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 201 201 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 356 356 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 365 365 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 398 398 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 621 621 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;

Query Match 39.7%; Score 2705.5; DB 1; Length 1167;
Best Local Similarity 46.2%; Pred. No. 6 1e-137;
Matches 563; Conservative 156; Mismatches 372; Indels 128; Gaps 24;

QY 48 ELYQGVQVQGNLELYLPTNASLFLQDIQVQGVYLIAHQVROVPLQRLIRIVRGTL 107
Db 51 KMYSGCNVYLENLEITYTQENQDLQFLQSIQVQGVYLIAHNEVSTIPLVNLRLIRGQL 110
QY 108 FEDNYALAVLDGDPINNTPTVTGASPGRLRELQRLSLTEILKGGVLIQRLPOLCYQDTI 167
Db 111 YEGNFTLLVMSYQK-NPSSP--DVYQVGLKQLQSLNTEILSGGVKVSHPNLLCNVETI 167
QY 168 LWKIDIFHKNNQLALTLIDTNRGRACHPCSPMCKGRSGESEDCCOSLTRTVCAGC-AR 226
Db 168 NWDIVDKTSNPTMNLIPAFERQCKQCDHGVCNSWAPGPHCKQFKLCAEQCNR 227
QY 227 CKGPLPTDCCHQCAAGCTGPKXSHDCLACLFHNSGICELHCPALVYNTDTFESMPNPE 286
Db 228 CRGPKPIDCNEHCAGGCTGPRATDCLACRDFNDGDKTCTPPPKIYDIVSHQVVDNPN 287
QY 287 GYTFGASCVTACPNYLSLTDVGSCTLVCLHNEQVTAEDGTQRCCKSPCARVCYGLG 346
Db 288 IKYTFGACVCKECPNSYVYTE-GACVRSACAGMLEVD-ENGKRSCKPCDGVCPKVCDDIG 345
QY 347 MEHLREVRAVTSANTQEFAGCKKIFGSLAFLPESDGDPAASNTAPLQPOLQVFELEBI 406
Db 346 IGSLSNTAVNSTNIRSFNSCTKINGDIILNNSFEGDPHYKIGTMDPEHLNLTIVTKEI 405
QY 407 TCYLYISAMPDLSPLSVFQNLQVIRGRILHNGAYS-LTQGLIGISWGLSLRLSLGSL 465
Db 406 TCYLVIMWMPENMTSLSVFQNLLEIRGTRTFSGRGSFVVVQVVRHLQWLGLSLKEVSAGN 465
QY 466 ALIHNTHLCLFVHTVPWDLFRNPHQALLHTANRDEDECVCEGLACHQLCARGHGWGP 525
Db 466 VILKNTLQRYANTINWRRLERSEDSQSEYDART-----ENQCTNNECEDSGCWGP 518
QY 526 TCVCNCSQFLRGQCEVEECVRLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCV 585
Db 519 TMCVSLHVDGRGRCVASCNLLQGBPREAQVDGRVCQCHQCLQVDTSLTCTGPGPANC 578
QY 586 ACAHYKDPFFCVARCPSPGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKCPAE 645

```

Db 579 KSAHFQDPQPCIPCPHGILGDCDTL-IMKYADKMGQCQCHQNCCTGCGSLGSGRGD 637
Qy 646 QRASPLTSIVSAGVGLLVVLGVVFGILIKRQOKKIRKVTMRLLQETBELVPLTPSGA 705
Db 638 -IVSHSLAVGLVSLGTLITVIVALLVLLRRRIK-RKRTIRCLLQEKELVPLTPSGQ 695
Qy 706 MPNOAQRILKETELRKVKVLSGAGCTVYKGIWIPDGENVKIPVALKVLRENTSPKANK 765
Db 696 APNOAFRIILKTEFKDKDRVLGSGAGTVYKGLWNPDPGENIRIPVALKVLRENTSPKANK 755
Qy 766 EILDEAYVAGVSPYVSRILGICLTSTVQLVTLQMPYCGCLLDHVRNENRGLASQDLNLN 825
Db 756 EVLDEAYVMSVDHPVHPCRLGICLTSAVQLVTLQMPYCGCLLDVYRQHERICQWLLN 815
Qy 826 CMQIAKMSYLEDLVRLVHRLDARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVP 885
Db 816 CVQIAKMSYLEDLVRLVHRLDARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVP 875
Qy 886 IKWMALESILRRRTHQSDVMSYGVTVWELMTFCAPKPYDGIIPAREIPDLLEKGERLP 945
Db 876 IKWMALESILQWTHQSDVMSYGVTVWELMTFCAPKPYDGIIPAREIPDLLEKGERLP 935
Qy 946 ICTIDVYMWKCMWIDSEKPRELVSFBSRMARDPQRFVWIQNEDLGPASPLDSTFY 1005
Db 936 ICTIEVYMWKCMWIDSEKPRELVSFBSRMARDPQRFVWIQNEDLGPASPLDSTFY 992
Qy 1006 RSLLEDMDGLVDABEYLVQOQFCFPAPGAGVHHRHSSSTRSGGDLTJGLEP 1065
Db 993 RSLLEDMDGLVDABEYLVQOQFCFPAPGAGVHHRHSSSTRSGGDLTJGLEP 1020
Qy 1066 SEEAAPSPLAPSGAGSDVFDGLGMAKGLQSLTHDPSPLORYSEDPV-PLPSET 1124
Db 1021 SEEAAPSPLAPSGAGSDVFDGLGMAKGLQSLTHDPSPLORYSEDPV-PLPSET 1053
Qy 1125 DGVYAPLTCSPQPVYVQVVRPQ-----PSPRE-----GPLP-AARPAGATLERAK 1171
Db 1054 DGH-----EYVNPQSGTSSRLSDIYNPNVEDLTDGWPVSLSSQEAENFSP 1103
Qy 1172 TLPSPKNGVNDVAFAGVENPBYLTPQGAAPQHPHPPAFNLYWDDPPERG 1231
Db 1104 YLNTNQSL---PLVSSGSDDDPY---QAG-----YQAAF-----LPQTG 1138
Qy 1232 APPSTFKGTPTAENPEYLG 1250
Db 1139 ALTNGMFLPAENLEYLG 1157

RESULT 9

ERB3 HUMAN
ID_ERB3_HUMAN STANDARD; PRT; 1342 AA.
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (C-erbB3) (tyrosine kinase-type cell surface receptor HER3).
GN ERB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2687875;
RA Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERB3, a third member of the
RT ERB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors."
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;

RA Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene."
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Placenta;
RX MEDLINE=93282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RL tyrosine kinase."
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC 1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTK.
CC 1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC 1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC 1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC 1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC 1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC 1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC 1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC 1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC 1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M29366; AAA35790.1; -;
CC EMBL; M34309; AAA35979.1; -;
CC EMBL; S61953; AAA26935.1; -;
CC PIR; A36223; A36223.
CC HSP; P11362; LFQK.
CC Genew; HGNC:3431; ERBB3.
CC MIM; 190151; -;
CC InterPro; IPR000494; EGFR_L_domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF01030; Recep_L_domain; 2.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 3.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
CC PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
CC Alternative splicing.
FT SIGNAL 1
FT CHAIN 20 1342
FT DOMAIN 20 643
FT TRANSMEM 644 664
FT DOMAIN 665 1342
FT DOMAIN 709 966
FT NP_BIND 715 723
FT BINDING 742 742
FT ACT_SITE 834 834
FT DISULFID 186 194
FT DISULFID 190 202
FT
FT POTENTIAL.
FT RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
FT EXTRACELLULAR (POTENTIAL).
FT POTENTIAL.
FT CYTOPLASMIC (POTENTIAL).
FT PROTEIN KINASE.
FT ATP (BY SIMILARITY).
FT BY SIMILARITY.
FT BY SIMILARITY.
FT BY SIMILARITY.

RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=96096535; PubMed=8522190;
RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
RT "Cloning of the rat ErbB3 CDNA and characterization of the
RL recombinant protein.";
RN Gene 165:279-284 (1995).
[2]
RP REVISIONS TO 85; 513 AND 565.
RA Hellyer N.J., Koland J.G.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE OF 922-1097 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659 (1997).
CC -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U29339; AAC28498.2; -;
CC EMBL; U52530; AAC3050.1; -;
CC HSP; P11362; IFCG.
CC InterPro; IPR000494; EGFR_L domain.
CC InterPro; IPR000719; Euk_Pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF01030; Recep_L domain; 2.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 5.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
CC PROSITE; PS00109; PROTEIN KINASE TYR; 1.
CC PROSITE; PS50011; PROTEIN KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
CC SIGNAL 1 19
CC CHAIN 20 1339
CC DOMAIN 20 643
CC TRANSMEM 644 662
CC DOMAIN 663 1339
CC DOMAIN 183 259
CC DOMAIN 707 964
CC NP_BIND 713 721
CC BINDING 740 740
CC ACT_SITE 832 832
CC DISULFID 186 194
CC DISULFID 190 202
CC DISULFID 210 218
CC DISULFID 214 226
CC DISULFID 227 235

FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 286 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 556 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1028 1028 L -> P (IN REF. 3).
SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F2402BFDFF1E CRC64;
Query Match 34.5%; Score 2350.5; DB 1; Length 1339;
Best Local Similarity 40.8%; Pred. No. 5.9e-118;
Matches 522; Conservative 170; Mismatches 419; Indels 167; Gaps 34;
QY 5 ALCRWGLLLALLPPGAASCTVFNFTVSFVLRVPSASHLELYGCGVVOGNLELY 64
DB 27 AVCP-GTFLNGSLVTGADNQYTL-----YKLYEKCSVMGNLSIVL 67
QY 65 LPTNASLSFDIQEYQGVYLIHAHNOVRQVPLQRLRIVRGTLQFEDNYALAVLDNGPLN 124
DB 68 TGNADLSFLQWREVTGYVIVAMNEFSLPLNLRVRCGYVDGKFAIFVM-----LN 122
QY 125 NTPVTGASPGGLREIQLRSLTEILKGGVLIQNPQLCYODTILKWDIFHNQNLATLI 184
DB 123 YNT-----NSSHALRQLKFTQLTEILSGVYIEKNDKLCMDTIDWRDIVRVR---GAEIV 175
QY 185 DTNRSRACHPCSPMKSGSCWGESSEDCOSLTETVCAGGC-ARCKGLPTDCCHEQCAAG 243
DB 176 VRNNGANCPCPCHEVCVGK-RCWGFGPDCCQLTKTTCAPQCNCRFCFNGPNQCDEGAGG 234
QY 244 CTGPKGSDCLACLFHNSGICELHCPALVTYNTDTFESMPENPGRYTFGASCVTACPNY 303
DB 235 CSGPQDTDCFACRRFNDSGACVPCRCPELVYVYKLTLEENPHTKYQYGGVGVASCPHNF 294
QY 304 LSTDVSGSCTLVCPHMQEVTAEQGTQCEKSKPCARVCYGLGMEHLREVRVTSANIQE 363
DB 295 V-VDTQFCVRACPPDKMEVD-KHGLKMCPCGGLCPKACEGTGSG--SRYQTVDSSNIDG 350
QY 364 FACKKIFGSLAFLPSFDGDPASNTAPLOEQLOVPETLEEITGYLYISAWPDSLPLS 423
DB 351 FVNCITKLGLNDFLITGLNVDPWHKIPALDPEKLVNFTVREITGYLNIOSWPPHMFNS 410
QY 424 VFQNLQVIRGRILHNGAYS-LTLQGLGISWGLRSLRSLGSLALIHNTLHLCFVHTVPV 482
DB 411 VFNLTITIGRSLYNRGFSLLIMKNLVNTSLGFRSLKEISAGRVYISANQQLCYHHSLNW 470
QY 483 DQIFRNPHQALLHTA-NRPDEVCVGEGLACHQICARGHCWGPGPTQCWCSQFLRGECV 541
DB 471 TLLRGPSEERLDIKYDRPLGCELAEGKVCDDLPCSSGGCGWPGPGQCLSRNYSREGVCV 530
QY 542 EECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEPAQDCVACAHYKDPPECVACRP 601
DB 531 THCNFLQGEPRFVHEAQCFSCHEPCLMPEGTSTCNGSGSDGSDACARCAHDFPHCVNSCP 590
QY 602 SGVKPDLISYMPIWKFPEDEGACQPCPINCTHSC--VDLDDKGPAPQARASPLTSIVSAV 659

Db 591 HGILG-AGGIYKYPQACNECRCHENCYQGCNGPBLQDCGAEVLMKSPHLVIAVT 648
Qy 660 GILVAVVGVVFGILIKRQOKIR-KYTMRRLLQETELVEPLTPSGAMPNQAOQRIKLT 718
Db 649 G-LAVILMILGSGFLYWRGRRIQNKAMRYLGRGESIFLDPDS-EKANKVLARIFKET 705
Qy 719 ELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVG 778
Db 706 ELRLKVLGSGVFGVTHKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVG 765
Qy 779 SPVYSLRLGICLSTVQLVTPYGLDHRVNRGRGLSGODLLNMCQIAKMSVLED 838
Db 766 HAHIVRLVGLCPGSSGLVQYPLGSLLDHVKHRETLPQLLLNMGVQIAKMYYLEE 825
Qy 839 VRLVHRDLAARNVLKSPNPKITDFGLARLLDIDETEYHADGKVPKIKWMALESILRRR 898
Db 826 HSMVHRDLARNVLKSPNPKITDFGLARLLDIDETEYHADGKVPKIKWMALESILRRR 895
Qy 899 FTHQSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMIMVKC 958
Db 886 YTHQSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMIMVKC 945
Qy 959 WMIDSECRPRFELVSEFMSWARDPQRFVIONEDLGPASPLDSTFYRSLLDDMGDLV 1018
Db 946 WMIDENIRPTFKELANEFTRMARDPPRYLVIKRAS-GFGTGP--PAAPSPVLTKEL--- 998
Qy 1019 DAEYLVPQQGFCFDPAPGAGGMVHRHRSSTSRGSGDLTLGLEPSEE----- 1068
Db 999 -QEALEPEL-----DLDLLEEEGLATSLGSL 1028
Qy 1069 -----EAPRSLPASEG-----AGSDVFDGLGMGAAGLOSLTHDPSPLQ 1110
Db 1029 SLPTGTLTRPRGSSLLSPSSGPMYMNQSSIGLGEACLDASVGLGREGQFSPISLH-PIPRG 1087
Qy 1111 RYSEDPVPLPSETDGVY----APL-----TC-----SPOPE-----YVQPDVRRQPP 1150
Db 1088 R-----PASESEGHVTSSEAELOKSVCHRSRSPRPRGDSAYHSQRHSLTPV 1140
Qy 1151 SPREGP-----LPAARPAAGATLERAKTLSP-GKGVW-----KDVFAFGGAVE 1192
Db 1141 TPLSPGLEREDGNGYMPDTHLRGASSREGTILSSVGLSVLTGTEEDD-----E 1192
Qy 1193 NPEYLTPOGGAAPQHPPP 1210
Db 1193 EYEMNRKRGSP-PRPP 1209

RESULT 11

EGFR DROME
ID EGFR DROME STANDARD; PRT; 1426 AA.
AC P0412; O61601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DE (Gurken receptor) (Torpedo protein) (Drosophila relative of ERBB).
GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC -Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
[1]
RN
RP SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX MEDLINE=94350209; PubMed=8070664;
RA Clifford R., Schupbach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals that several genetically defined classes of alleles cluster in subdomains of the receptor protein.";
RL Genetics 137:531-550(1994).
RN
RP REVISIONS.

RA Clifford R., Schupbach T.;
RL Submitted (MAR-1993) to the EMBL/GenBank/DBJ databases.
[3]
RN
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Livneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both hormone binding and kinase domains.";
RL Cell 40:599-607(1985).
[4]
RN
RP SEQUENCE FROM N.A. CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RX STRAIN=Oregon-R; TISSUE=Embryo;
RC MEDLINE=87002474; PubMed=3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the Drosophila EGF receptor homolog transcripts.";
RL Cell 46:1091-1101(1986).
[5]
RN
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor specification in wild-type, Ellipse, and null mutant Drosophila.";
RL Dev. Biol. 205:129-144(1999).
[6]
RN
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Ananidis P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., Adair J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D., Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Bonos P., Bertram B.P., Bhandari D., Bolshakov S., Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., Fogle C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
[7]
RN
RP SEQUENCE OF 959-1073 FROM N.A.
RC STRAIN=Daekwanyeong;
RX MEDLINE=85137938; PubMed=2983232;
RA Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
RT "A Drosophila genomic sequence with homology to human epidermal

RT growth factor receptor.";
 RL Nature 314:178-180(1985).
 RN [8]
 RP SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
 RP ANALYSIS.
 RX MEDLINE=92038942; PubMed=1936959;
 RA Raz E., Schejter E.D., Shilo B.Z.;
 RT "interallelic complementation among DER/flb alleles: implications for
 RT the mechanism of signal transduction by receptor-tyrosine kinases.";
 RL Genetics 129:191-201(1991).
 RN [9]
 RN REVIEW.
 RP MEDLINE=97248481; PubMed=9094709;
 RA Perrimon N., Perkins L.A.;
 RT "there must be 50 ways to rule the signal: the case of the Drosophila
 RT EGF receptor.";
 RL Cell 89:13-16(1997).
 CC -!- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
 CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
 CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
 CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
 CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
 CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
 CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOEROSA
 CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
 CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
 CC CUTICLE.
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -!- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
 CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
 CC PROTEIN.
 CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND
 CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
 CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
 CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
 CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
 CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
 CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
 CC AND THORACIC AND ABDOMINAL GANGLIA.
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF052754; AAC08536.1; -;
 DR EMBL; AF052753; AAC08536.1; JOINED.
 DR EMBL; AF052754; AAC08535.1; -;
 DR EMBL; AF052752; AAC08535.1; JOINED.
 DR EMBL; K03054; AAA51462.1; -;
 DR EMBL; K03417; AAA51460.1; -;
 DR EMBL; K03416; AAA50965.1; -;
 DR EMBL; K03418; AAA51461.1; -;
 DR EMBL; AF109077; AAD26134.1; -;
 DR EMBL; AF109078; AAD26132.1; -;
 DR EMBL; AF109082; AAD26132.1; JOINED.
 DR EMBL; AF109079; AAD26130.1; -;
 DR EMBL; AF109081; AAD26130.1; JOINED.
 DR EMBL; AF109079; AAD26131.1; -;
 DR EMBL; AF109083; AAD26131.1; JOINED.
 DR EMBL; AF109080; AAD26135.1; -;
 DR EMBL; AE003454; AAF46732.1; -;
 DR EMBL; X02293; CAA26157.1; -;
 DR EMBL; X78920; CAA55523.1; -;
 DR EMBL; X78918; CAA55521.1; -;

DR EMBL; X78919; CAA55522.1; -;
 DR PIR; A00640; GOFPE.
 DR HSP; P11362; IFGK.
 DR Flybase; FBgn0003731; Egfr.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase.1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 7.
 DR SMART; SM00219; TyrKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
 KW Developmental protein.
 FT SIGNAL 1 30 POTENTIAL.
 FT CHAIN 31 1426 EPIDERMAL GROWTH FACTOR RECEPTOR.
 FT DOMAIN 31 868 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 869 889 POTENTIAL.
 FT DOMAIN 890 1426 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 938 1198 PROTEIN KINASE.
 FT NP_BIND 944 952 ATP (BY SIMILARITY).
 FT BINDING 971 971 ATP (BY SIMILARITY).
 FT ACT_SITE 1063 1063 BY SIMILARITY.
 FT MOD_RES 902 902 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
 Query Match 28.1%; Score 1931; DB 1; Length 1426;
 Best Local Similarity 32.8%; Pred. No. 1.4e-35; Indels 342; Gaps 42;
 Matches 464; Conservative 180; Mismatches 459;
 QY 37 LRVPKVSASHL-----ELYQCQCVQVQGNLELTLYPT-NASLSFLQDIOEQVGYVLIHQV 91
 DB 109 LSVPSNKEHHYRLRDORYTCTVDGNLKLTLWLPNENLDLSFLDNIREVTGYILSHVDV 168
 QY 92 RQVPLQRLRIVGTQLF-----EDNVALAVLDNGDPLNNTPTVTGASPGGLRELQKLSLT 146
 DB 169 KKVVFVKLQIRGRTLFLSLVSEBEKALFV-----TYSKMYLTLEIFDLR 212
 QY 147 EILKGGVLIQNPQLCVQDTILWKDIFHKNNQLALTIDNRSRACHPCSPMKSGSCWG 206
 DB 213 DVLNGQVGFHNNYLNLCMRITQMSVNSGTDAYNDFTAPERECPKCHESCTHG-CWG 271
 QY 207 ESSEDQCSLRTRVCAGGCA--RCKGPLPTDCHEQCAAGCTGPKHSDCLACLHFNHSGIC 264
 DB 272 EGPKNCKFSKLTCSQCAGRCVCGPKPRECCHLFCAGGCTGPTQKDCIACKNPFDEAVS 331
 QY 265 ELHCPALVYNTDTFESMPNPEGRYTFGASCVTACPNYLSLTDVGSCTLVCPHLNQEVTA 324
 DB 332 KECCPPMRKYNPTTYVLETNPEGKYAYGATCKVECP-GHLRLDNGACVRSQPDQMDKGG 390
 QY 325 EDGTQCEKCSKPCARVCYGLGMEHLREVPVTSANIQEFAGCKKIFGSLAFIPESFDG- 383
 DB 391 E-----CVPNGPCPKTCPGTVLH-----AGNIDSPNCTVIDGNIRILDQTSFGF 437
 QY 384 -DPASNTA-----PLQPEQLQVFETLEITGYLYTISAMPDLSLPDLSVFONLQVIRGIL 436
 DB 438 QDVYANYTMGPRVPIPLDPERREVFSTVKETGYINIEGTHPQFNLVSFRNLETHGRQL 497
 QY 437 HNGAY-SLTLOGLGISWLGLRSRLRELSGSLALIHNNTHLCFVHTVPMQDLFRPNQALLH 495
 DB 498 MESMFAALAIVKSSLYSLEMRNLKQISSGSWVIQHNRLDLCVSNIRNPAIQKEQKVVV 557
 QY 496 TANRPDECVGEGLACHQLCARGHCWGPGPTQCVCNCSQFLRGQBCVCEECRVLQGLPREYV 555
 DB 558 NENLRADLCENKGTICSDQCNEDGCWGAGTDQCLTKNFNGTCTIADCGVISNAYK--F 615
 QY 556 NARHCLPCHPECPQNSVTCFGEADQCVCACAHYKDPFFCVARCP----- 601

```
Db 616 DNRTCKIKCHPECR-----TCNGAGADHCQECVHVRDQHCVCSECPKKNKYNDRGVCRECH 669
Qy 602 -----SGVK-----PDLSTMPYWKPF--PDEEG 621
Db 670 ATDCGTGPKDTIGAGACTTCNLAIINNDATVKRCLLKDDKCPD-GY--FWYVHQPQEQG 726
Qy 622 ACQP-----CPI-----NCTH-----632
Db 727 SLKPLAGRAVCRKCHPLCELCCTNGYHEQVCCKTHYKRREQCETECPADHYTDEQREC 786
Qy 633 -----SC-----VLDLDDKG-----CPAQOR-----647
Db 787 FORHPECNGCTGPGADCKSRNFKLFDANETGPNVSTWNTFNCTSKCPLMRHNVYQYTA 846
Qy 648 -----ASPLTS-----IVSAVVGILLVVVLGVVFGILIKRQOQKIRKYT--686
Db 847 IGPYCAASPRSSKITANLDVNMFIITGAVLPTICILCV--TYICROKQAKETVK 904
Qy 687 MRRLLOTELVLEPTPGAMPNQAMRILKETELRKVKVLGSGAFGVYKGIWIPDGENV 746
Db 905 MTWALSCEDESEPLRPSNIGANCKLRIVKDAELRKGVLGGMGAFGRVYKGVVWPEGENV 964
Qy 747 KIPVAIKVLENTSPRANKELDEAYVMAGVGPVYSRLLIGICLTSTVQLVTLMPYGCIL 806
Db 965 KIPVAIKELLUKSTGAESSEEFLEAYIMASEEHVNLKLLAVCMSSQMMILITQLMPLGCL 1024
Qy 807 LDHVRENRGRGLSGDILLNWCQAKGMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGL 866
Db 1025 LDVVRNRDKIGSKALLNWSQTAKGMSYLEEKLVRDLAARNVLVOTSLVKITDFGL 1084
Qy 867 ARLLDIDETEHADGGKVPKIKWMALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGI 926
Db 1085 AKLLSSDSNEYKAAGGKMPKWLALCEIRNRVFTSKSDVWAFGVTIWELLTFQGRPHENI 1144
Qy 927 PAREIPDLLEKGERLQPPCTTIDVVMVWKWIMIDSECPRPREFLVSEFSRWARDPQRF 986
Db 1145 PAKDIPDLIEVGLKLEQPEICSLDIYCTLLSCHWLDAAARPTFKQLTTFVFAEFARDPGRY 1204
Qy 987 VTIQNEIDL--PASPLDSTFYRSLLEDD--DMGDLVDAEYLVPOQGFPCPPAPGAGG 1041
Db 1205 LAIPGDKFTLPA-----YTQDEKDLRLKLAPTTDGSEAIKAPDDYIQPKAAACPS- 1256
Qy 1042 MVHHRHSSTSGGDLTLGLFSEEEAP-----RSPLAPSEGAGSDVFDG---DLGM 1092
Db 1257 -----HRTDCT-----DEMPKLNRYCKDPSNKNSSGTGDDERSSAREVG 1296
Qy 1093 GAAKGLQSLTHDPSPLORYSEDPVLPSETDGYVAPLTCSPQPEYVNPQDVRPOPSP 1152
Db 1297 GNLR-----LDLPVDEDDYLMPTCPQGPNNNNNMN-----NP 1328
Qy 1153 REGPLPAARPAATLERAKTSLSPKNGVVKDVFAGGAVENPEYL---TFQCGGAAPOPH 1208
Db 1329 NQNNMAAVGVAAGYM-----DLGVPVSVDNPNPYLLNAQTLGVGESPIPT 1373
Qy 1209 -----PPAFSP-AFDNLYTYMD 1224
Db 1374 QTIGIPWGGPGTMEVKVPMGPSEPTSSDHEYND 1408
```

RESULT 12

```
ERBB_ALV
ID ERBB_ALV STANDARD; PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (BC 2.7.1.112).
GN V-ERBB.
OS Avian leukosis virus.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
ON NCBI_TaxID=11864;
RX [1]
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nilssen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA Crittenden L.B., Raines M.A., Kung H.-J.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor.";
RL Cell 41:719-726(1985).
CC -|- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -|- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -|- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -|- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M10066; AAA48763.1; ALT_INIT.
DR PIR; A00643; TVCHLV.
DR FIC; B00643; TVFVLV.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000715; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT SITE 257 257 BY SIMILARITY.
FT ACT SITE 257 257 BY SIMILARITY.
SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;

Query Match 25.7%; Score 1749.5; DB 1; Length 634;
Best Local Similarity 52.3%; Pred. No. 2.7e-86;
Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17;

QY 587 CAHYKDPPCVACRCSGVKPDLSYMPVWPFDEGACQPCPINCTHSCVDLDDKGGPAEQ 646
Db 3 CAHFIDGPRCHVACAPAGLVGENDTL-VWKYADANAVCOLCHFNCTRGCGPGLGCP--- 58
QY 647 RASPLTSIVSAVY-GILLVVVLGVVFGILIKRQOQKIRKYTMRLLOETELVEPLTPSGA 705
Db 59 NGSKTPSIAAGVVGGLLCLVVVVGGLGIGLYLRR-HIVKRLRLRLQERLEPLTPSGE 117
QY 706 MPNQAMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLENTSPKANK 765
Db 118 APNQAHRLILKETEFKVKVVLGSGAFGVYKGIWIPDEGEKVKIPVAIKELREATSPKANK 177
QY 766 EILDYAVWAGVSGPVSRLLIGICLTSTVQLVTLMPYGCILDHVRENRGRGLSQDLN 825
Db 178 EILDYAVWAGVSGPVSRLLIGICLTSTVQLVTLMPYGCILDIYREHKDNGISQYLLN 237
QY 826 CMOIAGMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP 885
Db 238 CVOIAGMSYLEERLLVHRDLAARNVLKTPQHKITDFGLAKLLGADKEYHAEGKVP 297
QY 886 IKWMALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPP 945
```



```
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGI PASEISSVLEKGERLPQP 357
Qy 946 ICTIDVYIMVKCMWIDSECRPRRELVSFSEMRDQRFVVIQ-NEDLGPASPLDSTF 1004
Db 358 ICTIDVYIMVKCMWIDSECRPRRELVSFSEMRDQRFVVIQ-QGDERMHLPSPTDSKF 417
Qy 1005 YRSLEDDDDGDLVDAEYLVPOQGFPCPDAPGAGGGMVHRRSSSTRSGGDLTLGLE 1064
Db 418 YRLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1065 PSEEEAPRSL-----APSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDTVP 1119
Db 450 -----SRTPLLSLSATSNNSATNCID-----RNGQGHVPVREDSFQRYSSDPTGN 495
Qy 1120 LPSET--DGVVAPLTCSPOPEYVNDVVRPQPPSPREGPLPAARPAGATLERAKTLPSPGK 1177
Db 496 FLEESIDDDGL-----PAPEYVQ--LMPKPKS----- 524
Qy 1178 NGVVKDVF-----AFGGAIVENPEVLTPOGGAAPQHPHPPAFSPAFDNLXY 1222
Db 527 NQIYNNISLTAISKLPMDSRQNSHSTAVDNPEYL-----NTNOSPLAKTVFESSPY 578
Qy 1223 WDQ-----DPPE-----RGAPPSTFGTGTPTAENPEYLGIDVP 1254
Db 579 WIQSGNHQINLDNPDYQDFLFPNETKPNGLLKVPAAENPEYLRVAAP 625

RESULT 13
ERBB AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUN-1986 (Rel. 01, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
RT gene family."
RL Cell 35:71-78(1983).
RN [2]
RX SEQUENCE OF 1-152 FROM N.A.
RP MEDLINE=84223957; PubMed=6328658;
RA Debouire B., Henry C., Benaisa M., Biserte G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erbA gene of avian erythroblastosis virus reveals a
RT new type of oncogene."
RL Science 224:1456-1459(1984).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
CC -1- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
DR EMBL; K02006; AAA42394.1; ALT_INIT.
DR EMBL; K01216; AAA42400.1; -.
DR PIR; A00644; TVYUH.
DR HSSP; P11362; LFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
KW DOMAIN 132 399 PROTEIN_KINASE
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 145 145 I -> V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MW; 76BCDD06745D609 CRC64;

Query Match 25.0%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 7.6e-84;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

Qy 587 CAHYKPPFCVACPSGVKPDLSYMPIWKEPDEGACQPCINCTHSCVDLDDKGCPEAQ 646
Db 3 CAHFIDGPHCVKACPAVLGENDTL-VRYADANAVCOLCHPNCTRGCKPGLEGCP--- 58
Qy 647 RASPLTSIVSAVV-GILLVVLGVWFGILIKRQOKIRKVTMRLLQETVELVPLTPSGA 705
Db 59 NGSKTFSIAAGVVGGLLVVGLGIGLYLRR-HIVRKTLLRLLQERLEFLTPSGE 117
Qy 706 MPNQAMRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPKANK 765
Db 118 APNOAHLRLKETEFKKVKVLGSGAGFTTYKGLWIPEGEKVKIPVAIKELREATSPKANK 177
Qy 766 EILDEAYVWAGVSPVYRLLGLTSTVQLTQVLPYGLLDHVRNRLGSDLLNW 825
Db 178 EILDEAYVWASVDNPHVCRLLGLTSTVQLTQVLPYGLLDYIREHKDNISQYLLNW 237
Qy 826 CMQIAKMSYLEVRLVRLDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVP 885
Db 238 CVQIAGKMYLEERLVRHDLAARNLVKTPQHVKITDFGLAKLGADEKEYHAEGKVP 297
Qy 886 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQP 945
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGI PASEISSVLEKGERLPQP 357
Qy 946 ICTIDVYIMVKCMWIDSECRPRRELVSFSEMRDQRFVVIQ-NEDLGPASPLDSTF 1004
Db 358 ICTIDVYIMVKCMWIDSECRPRRELVSFSEMRDQRFVVIQ-QGDERMHLPSPTDSKF 417
Qy 1005 YRSLEDDDDGDLVDAEYLVPOQGFPCPDAPGAGGGMVHRRSSSTRSGGDLTLGLE 1064
Db 418 YRLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1065 PSEEEAPRSL-----APSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDTVP 1119
Db 450 -----SRTPLLSLSATSNNSATNCID-----RNGQGHVPVREDSFQRYSSDPTGN 495
Qy 1120 LPSET--DGVVAPLTCSPOPEYVNDVVRPQPPSPREGPLPAARPAGATLERAKTLPSPGK 1177
Db 496 FLEESIDDDGL-----PAPEYVQ--LMPKPKS----- 524
Qy 1178 NGVVKDVF-----AFGGAIVENPEVLTPOGGAAPQHPHPPAFSPAFDNLXY 1222
Db 525 --VQNIYFISLTALSKLPMDSRQNSHSTAVDNPEYL-----NTNOSPLAKTVFE 574
Qy 1219 NLYYWDQPPERGAFFPTFGTPTAENPEY 1248
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.3575 Seconds
(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-5-25-14
Perfect score: 6812
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues 671580
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL 21:.*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_rvirus:*
 - 16: sp_bacteriap:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	6183	90.8	1259	6	O18735	O18735 canis famil
2	3140	46.1	1209	11	Q9QX70	Q9qx70 rattus norv
3	3109	45.6	1210	11	Q9EP98	Q9ep98 mus musculu
4	2739	40.2	1165	13	Q9YH40	Q9yh40 xiphophorus
5	2729.5	40.1	1137	13	Q9W6F6	Q9w6f6 gallus gall
6	2292	33.6	1328	13	P79754	P79754 fugu rubrip
7	2000.5	29.4	1433	5	Q9BIH9	Q9bih9 anopheles g
8	1782.5	25.9	419	4	Q9UK79	Q9uk79 homo sapien
9	1739	25.5	367	11	Q8R2X1	Q8r2x1 mus musculu
10	1720	25.2	729	15	Q86712	Q86712 avian rous-
11	1718	25.2	567	15	Q86714	Q86714 avian rous-
12	1697.5	24.9	412	4	Q8WYV0	Q8wyv0 homo sapien
13	1653.5	24.3	962	15	Q64895	Q64895 avian eryth
14	1645	24.1	545	15	Q85468	Q85468 avian eryth
15	1500.5	22.0	655	11	Q9WVF5	Q9wvf5 mus musculu
16	1484.5	21.8	643	11	Q9ERV6	Q9erv6 mus musculu

17	1275	18.7	1193	5	Q9YIX8	Q9yix8 ephydatia f
18	1197.5	17.6	1368	5	Q23821	Q23821 caenorhabdi
19	1175	17.2	1717	5	Q26566	Q26566 schistosoma
20	1118	16.4	527	13	Q90836	Q90836 gallus gall
21	995.5	14.6	478	11	Q9ESE0	Q9ese0 rattus norv
22	934.5	13.7	599	13	Q9PSH2	Q9psh2 gallus gall
23	906	13.3	165	4	Q14256	Q14256 homo sapien
24	887	13.0	176	11	Q923V5	Q923v5 rattus norv
25	806.5	11.8	346	13	P11776	P11776 xiphophorus
26	778	11.4	435	5	Q8SZW1	Q8szw1 drosophila
27	756	11.1	1362	13	Q9PVZ4	Q9pvz4 xenopus lae
28	754.5	11.1	311	13	Q99162	Q99162 xiphophorus
29	735.5	10.8	1671	5	Q9NUV5	Q9njv5 biophalar
30	723	10.6	149	6	Q9BG66	Q9bg66 oryctolagus
31	710.5	10.4	331	4	Q9BUD7	Q9bud7 homo sapien
32	701.5	10.3	1368	13	Q8UW85	Q8uw85 paralichthy
33	696	10.2	1418	13	Q93457	Q93457 scopthalmu
34	679.5	10.0	1369	13	Q8UW86	Q8uw86 paralichthy
35	670	9.8	1358	13	Q73798	Q73798 xenopus lae
36	659	9.7	1412	13	Q8UW84	Q8uw84 paralichthy
37	657.5	9.7	1472	5	Q9U5A8	Q9u5a8 bombyx mori
38	645	9.5	1245	13	Q9YGH8	Q9ygh8 scopthalmu
39	642.5	9.4	1418	13	Q8UW83	Q8uw83 paralichthy
40	634	9.3	1371	11	Q9QVW4	Q9qvw4 rattus sp.
41	626	9.2	2144	5	Q9VD94	Q9vd94 drosophila
42	605	8.9	935	4	Q96L35	Q96l35 homo sapien
43	601	8.8	987	11	Q91YMO	Q91ymo mus musculu
44	598	8.8	987	11	Q99MR2	Q99mr2 mus musculu
45	587.5	8.6	1036	4	Q07912	Q07912 homo sapien

ALIGNMENTS

RESULT 1

O18735 O18735 PRELIMINARY; PRT; 1259 AA.
ID O18735
AC O18735;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Erbb-2.
OS Canis familiaris (dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT "cDNA cloning of erbb-2 from canine mammary gland.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB008451; BAA23127.1; -
DR HSSP; P11362; IFGK.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; Yfp_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP_2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE; PS0107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;

Query Match		90.8%; Score 6183; DB 6; Length 1259;
Best Local Similarity		90.5%; Pred. No. 0;
Matches 1144; Conservative		43; Mismatches 63; Indels 14; Gaps 4;
Qy	1	MELAAACRWGLLLALLPAGAASTOVCTFNNTVSWFLRVKVSASHLE----LYOGCOVV 56
Db	1	MELAAACRWGLLLALLPAGAASTOVCTFNNTVSWFLRVKVSASHLE----LYOGCOVV 56
Qy	57	QGNLELYTPNASLSFLQDIQEVQGYVLIHAHNOVROVLPQRLRIVRGTOLEFEDNYALAV 116
Db	57	QGNLELYTPNASLSFLQDIQEVQGYVLIHAHNOVROVLPQRLRIVRGTOLEFEDNYALAV 116
Qy	117	LDNGDPLNNTTTPVTGASPGGLRELQLRLSLTEILKGGVLIQRNPOLCYQDTILMKDIFHN 176
Db	117	LDNGDPLEGGIPAPGAAGGLRELQLRLSLTEILKGGVLIQRNPOLCHQDTILMKDVFHN 176
Qy	177	NQLALTILIDNRSRACHPCSPMGKRCWGBSSSDCQSLTRTVCAAGCARKGKPLPTDCC 236
Db	177	NQLALTILIDNRSRACHPCSPMGKRCWGBSSSDCQSLTRTVCAAGCARKGKPGPTDCC 236
Qy	237	HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
Db	237	HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
Qy	297	TACPNYLTSTDVGSCITLVCPHNOQVTAEDGTQRCCKSPCARVCYGLGMEHLREVRV 356
Db	297	TSCPNYLTSTDVGSCITLVCPHNOQVTAEDGTQRCCKSPCARVCYGLGMEHLREVRV 356
Qy	357	TSANIOBAGCKITFGLAFIPESFDGDPASNTAPLOEQLOVETLEETGYLYISAMP 416
Db	357	TSANIOBAGCKITFGLAFIPESFDGDPASNTAPLOEQLOVETLEETGYLYISAMP 416
Qy	417	DSLPLSVFONQVIRGRLHNGAYSLTLOGLSLWGLSLRGLSGLALIHNTLCP 476
Db	417	DSLPLSVFONQVIRGRLHNGAYSLTLOGLSLWGLSLRGLSGLALIHNTLCP 476
Qy	477	VHTVPDQLFRNPQALLHTANRDEBCVGBGLACHOLCARGHGWGPGTQCVNCSQFLR 536
Db	477	VHTVPDQLFRNPQALLHTANRDEBCVGBGLACHOLCARGHGWGPGTQCVNCSQFLR 536
Qy	537	GOECVEGRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQCACAHYKDPFFC 596
Db	536	GOECVEGRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQCACAHYKDPFFC 595
Qy	597	VARPCSGVKPDLVMPYTWKPDDEGACQPCPINTCHSCVDLDDKCAEQASPLTIVS 656
Db	596	VARPCSGVKPDLVMPYTWKPDDEGACQPCPINTCHSCVDLDDKCAEQASPLTIVS 655
Qy	657	AVGILLVVLGVVFGILIKRRQOKIRKYMRLLOETELVEPLTPSGAMPNQAQMRILK 716
Db	656	AVGILLVVLGVVFGILIKRRQOKIRKYMRLLOETELVEPLTPSGAMPNQAQMRILK 715
Qy	717	ETELRKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRENTSPKANKILDEAYVMAG 776
Db	716	ETELRKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRENTSPKANKILDEAYVMAG 775
Qy	777	VGSYPVSRLLGICLTSTVOLVTLMPYCYLLDHHVRENRLGSDLLNWCQIAKGSYL 836
Db	776	VGSYPVSRLLGICLTSTVOLVTLMPYCYLLDHHVRENRLGSDLLNWCQIAKGSYL 835
Qy	837	EDVRLVHRDLAARNVLKSPNHVKITDGLARLLDIDETEYHADGGKVPKMWALLESILR 896
Db	836	EDVRLVHRDLAARNVLKSPNHVKITDGLARLLDIDETEYHADGGKVPKMWALLESIPP 895
Qy	897	RRFTHQSDVWSGYVTVMELTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYIMV 956
Db	896	RRFTHQSDVWSGYVTVMELTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYIMV 955
Qy	957	KCMWIDSECRPRFRELVSSEARMARDPQRFVIONEDLGPASPLDSTFYSLLEDDDDMGD 1016
Db	956	KCMWIDSECRPRFRELVAEFSARMARDPQRFVIONEDLGPASPLDSTFYSLLEDDDDMGD 1015
Qy	1017	LVDAAEYLVPOQGFPCPDPAFGAGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSPLA 1076

Db	1016	LVDAAEYLVPOQGFPCPDPAFGAGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSPLA 1075
Qy	1077	PSGAGSDVFDGDLGMAAGKLOSLTHDPSPLORYSEDTVPLPSETDGYVAPLTCSPQ 1136
Db	1076	PSGAGSDVFDGDLGMAAGKLOSLTHDPSPLORYSEDTVPLPSETDGYVAPLTCSPQ 1135
Qy	1137	PEVNOPDVVRPQPPSPREGPLPAARPAAGATLER-----AKTLSPGKNGVVVKDVFAGGAV 1191
Db	1136	PEVNOPDEWVPQPPSPREGPLPAARPAAGATLERPKTLSPKTLSPGKNGVVVKDVFAGGAV 1195
Qy	1192	ENPEYLTPOGGAAPQHPPPAFSPAFDNLYWQDPPERGAAPPSTFKGTPTAENPEYGL 1251
Db	1196	ENPEYLAAPRAAPQHPPPAFSPAFDNLYWQDPPERGAAPPSTFKGTPTAENPEYGL 1255
Qy	1252	DVPV 1255
Db	1256	DVPV 1259
RESULT 2		
Q9QX70	PRELIMINARY;	PRT; 1209 AA.
AC	Q9QX70;	
DT	01-MAY-2000 (T-EMBLrel. 13, Created)	
DT	01-MAY-2000 (T-EMBLrel. 13, Last sequence update)	
DT	01-JUN-2002 (T-EMBLrel. 21, Last annotation update)	
DE	Epidermal growth factor receptor.	
GN	EGFR.	
OS	Rattus norvegicus (Rat).	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.	
OX	NCBI_TaxID=10116;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
RX	MEDLINE=90258888; PubMed=2342466;	
RA	Petch L.A.; Harris J.; Raymond V.W.; Blaeband A.J.; Lee D.C.;	
RA	Earp H.S.;	
RT	"A truncated, secreted form of the epidermal growth factor receptor is	
RT	encoded by an alternatively spliced transcript in normal rat tissue."	
RL	Mol. Cell. Biol. 10:2973-2982(1990).	
RP	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
RA	Petch L.A.;	
RL	Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
RA	Guttridge K.; Dawson T.L.; Earp H.S.;	
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; M37394; AAF14008.1; -.	
DR	HSP; P11362; 1FGK.	
DR	InterPro; IPR000494; EGFR L domain.	
DR	InterPro; IPR000719; Euk_pkinase.	
DR	InterPro; IPR002174; Furin-like.	
DR	InterPro; IPR001245; Tyr_kinase.	
DR	Pfam; PF00057; Furin-like; 1.	
DR	Pfam; PF00069; pkinase; 1.	
DR	Pfam; PF01030; Recep L domain; 2.	
DR	PRINTS; PR00109; TYRKINASE.	
DR	ProDom; PD000001; Euk_pkinase; 1.	
DR	SMART; SM00261; FU; 3.	
DR	SMART; SM00219; TyrKc; 1.	
DR	PROSITE; PS00107; PROTEIN KINASE ATP; 1.	
DR	PROSITE; PS00011; PROTEIN KINASE DOM; 1.	
DR	PROSITE; PS00109; PROTEIN KINASE TYR; 1.	
KW	ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.	
SQ	SEQUENCE 1209 AA; 134891 MW; 96FEF7F6CC1B773 CRC64;	
Query Match		46.1%; Score 3140; DB 11; Length 1209;
Best Local Similarity		49.9%; Pred. No. 1e-226;

[illegible]

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00101; PROTEIN_KINASE_DOM; 2.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1165 AA; 129614 MW; 7F7EE38D8771A74E CRC64;

Query Match
Best Local Similarity 40.2%; Score 2739; DB 13; Length 1165;
Matches 586; Conservative 162; Mismatches 380; Indels 148; Gaps 31;

QY 1 MELAALCRWGLLALLPPG-AAST- -TFNNFTV--SFMLRVPKVSASHLELYQ 51
DB 4 LELLELE- -LLLLLLSIGRCCSTDPDRKVCQGTSNQMTLMDNHLKMKK-----MYS 52

QY 52 GCQVQGNLELYLPTNASISFLQDIQEVGYVLIHNRQVPLQRLIRVGTQLPEDN 111
DB 53 GCNVLENLEITYTQENQDLNFLQSIQEVGYVLIAMNEVSTIPLVNLRLIRGQNLVEGN 112

QY 112 YALAVLDNGDPLNNTTPTVTCASPGGLRELQRLSLTEILKGVLIQRNPOLCYQDTILWKD 171
DB 113 FTLLVWSNYQK-NPSSP- -DVYQVGLKQLQLSNLTELISGGVKVSHNPLLCNVETINWMD 169

QY 172 IFHKNQQLALTLIDTNSRACHPCSPKCGSRGSESDCOSLTRTVACAGGC-ARCKGP 230
DB 170 IVDKTSNPTWNLPHAFERQCQKCDPCVNGSCWAPGHCQKFTKLLCAEQNRCRGP 229

QY 231 LPTDCHEQCAAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 290
DB 230 KPIDCCNEHCAGGCTGPRATDCLACRDFNDGDKTCTPPKIYDIIVSHQVNDPNIKYT 289

QY 291 FGASCVTACPNYLSLTVGSGTLCVPLHNOEVTAEADGTQRCCKSKPCARVCGLGMHEHL 350
DB 290 FGAACVKECESNYVTE-GACVRSKCSAGMLEVD-ENGRSKCKPDGVCPRKVCIDGIGISL 347

QY 351 REVRVAVTSANIOEAGCKIFGSLAFLEPSFGDGPASNTAPLOEQQLQVETLEELITGYL 410
DB 348 SNTIADVNTWIGFSNCTKINGDIILNRNSFEGDPHYKIGPMOPEHLWNUTTKKEITGYL 407

QY 411 YISAWPDSLPDLSVFQNLQVIRGILHNGAYS-LTLQGLGISWLGSLRSLRELASGLALIH 469
DB 408 VIMWPNEMTSLVSFQNLLEIRGRTTFSRGFSFVVVQVSHLQWLGSLLSKEVSAGNVILK 467

QY 470 HNTHLCFVHTVPWDQLFRNPHQALLHTANRPEBECVGEGLACHQLCARGHCWPGPTQCV 529
DB 468 NTPQLRYASTINWRRLFRSEDQSTIEYDART- -ENQTCNCESEDGCGWPGPTMCV 520

QY 530 NCSQFLRGQCEVECRVLQGLPREYVNAHCLPCHPEQCQNGSVTCFCGEAQCACAH 589
DB 521 SCLHVDGRGRCVASCNLLQGEPREAQVDGRVCVQCHQCEQLVQTDLSLTCYGFPGPANCSCAH 580

QY 590 YKDPFFCVARCPGSKPDLSPYMPIWKFPDBEGACQPCPINCTHSCVDLDDKGCPAEQRAS 649
DB 581 FQDGPQICPRCPHGLMGDGL-LWKYADKMGQCPCHQCTQCGSPGLSGCGED-IVS 638

QY 650 PLTSIVSAVGVILLVVLGVVFGILLIKRROOKIRKTYMRLLQTELVEPLTSGAMPNQ 709
DB 639 HSSLAVGLVSGLLITVALLIVLLRRRIK-RKRTIRLLQEKELVEPLTSGQAPNQ 697

QY 710 AQMRILKETELRKVKVLGSGAGCTVYKGIWIPDGENVKIPIVAILVRENTSPKANKEILD 769
DB 698 AFURLKETEFKDRVLGSGAGFVYKGLWNPGENIRIPVAILVRENTSPKANKEILD 757

QY 770 EAYVMAGVGPVYSRLIGICLTSTVQLVTLQMPYGCLLDHVRENRLGSLQDILLNWCQI 829
DB 758 EAYVMASVDHPHVCRLIGICLTSAVQLVTLQMPYGCLLDVVRQHERICQWLLNWCQI 817

QY 830 AKGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHYHADGKGVPIKWM 889
DB 818 AKGMNYLEERHLVHRDLAARNVLLKNPNVKITDFGLSKLLTADKEYQAHGKGVPIKWM 877

QY 890 ALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTI 949
DB 878 ALESILQWITYTHQSDVMSYGVTVWELMTFGSKPYDGIIPAKEIASVLENGERLPQPICTI 937

Query Match 40.1%; Score 2729.5; DB 13; Length 1137;

QY 950 DVTMIMVKCMIDSECRPRFRELVSFBSRMARDPQRFVVIQNEDLGAPSLDSTFYRSLL 1009
DB 938 EVYMIILKCMIDPSSRRPRFRELVGFSQWARPDSRYLVIQG---NLSPSPDRRLFSL 994

QY 1010 EDDMDGLVDAEYLVPOQGFPCDPAPGAGGVMVHRHSSSTRSGGDLTLGLEPSEEE 1069
DB 995 SSDSD--DVVDAEYLL- - - - -RYKRIN-RQGS----- 1018

QY 1070 APRSLAPSEAGSDVFDGDLGMAAKGLQLSLTHDPSPQLQRYSEDPV-PLPSETDGYV 1128
DB 1019 ---EPCIPNGH-----PVRENSIALRYISDPTQNALEKDLGDH- 1054

QY 1129 APLTCSQPEYVNPQVDRPQP-----PSPRE-----GPLP-AAPAGATLPRAKTLSP 1175
DB 1055 -----EYVNPQGETSSRLSDIYNPNYEDLTDCWGPVSLSSQEAETNFSRPEYLVNT 1105

QY 1176 GKNGVVKDVFAFGCAVENPEYLTPOGGAAPQHPHFAFPAFDNLYWDDPDPERGAPPS 1235
DB 1106 NQNSL---PLVSSGSMDDPDY---QAG-----YQAAF-----LPQTGALTG 1140

QY 1236 TFKGTPTAENPEYVGL 1251
DB 1141 NGMFLPAEENLEYLGL 1156

RESULT 5
Q9W6F6 PRELIMINARY; PRT; 1137 AA.

AC Q9W6F6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Receptor tyrosine kinase (Fragment).
GN EBB4.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HINDBRAIN;
RX MEDLINE=99263203; PubMed=10328884;
RA Dixon M., Lumsden A.;
RT "Distribution of neurokinin-1 (nrg1) and erbB4 transcripts in embryonic chick hindbrain";
RL Mol. Cell. Neurosci. 13:237-258 (1999).
DR EMBL; AF121963; AAD31764.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001368; TNFR_c6.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recept_L domain; 1.
DR Pfam; PF02757; YLP; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.
KW Kinase; Tyrosine-protein kinase.
FT NON_TER 1
SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC84F CRC64;

QY 408 GYLIIISAWPSLDLSVFNQLOVIRGRILHNGAVSLTQGLGSLWGLRLSLRELGLSLAL 467
Db 391 DILNIQSWPKELNDLSVSSLTITQGRSLFKRFSLMVMIRPTLTSLGLRLSLREISDGSVY 450
QY 468 IHNHNLCLFVHTVPDOLFRNPH-QALLHTANRPEDECVGEGACHOLCARGHCWGPPT 526
Db 451 ISQNAHLCYHTVNWTOGLFRGSRVANSLSNRPMACVADGRVCDPLCSDSCGCGPDP 510
QY 527 QCVNCSOFLRGQECVEECRVLQGLPREYVNAH-CLPCHPECPQNGSVTCFQPEADQCV 585
Db 511 QCLSCRYSRHGTCAVCHFNISGPIREFAGLNGVCVACHPECKPQTCKASCTGPGADECM 570
QY 586 ACAHYKDPPECVACRCPGVDPDLSYMPIWKFPEDEGACQCPNCTHSCVDLDDKGPAP 645
Db 571 ACTFRDGPCTCMSCGPAVN-DGEKGLIFKPPNREHCEFCQHCQCTQCGSGPGLNDC--- 626
QY 646 QRASPLTSIVSAVVGILLVVVLGVF-----GILIKRQKQKRYTMRLLQETELVE 698
Db 627 LEAARLTISSQITGIALGVPAGLIFCLVLFGLQMLVHRGLAIRKRAMRYLESSESFE 686
QY 699 PLTPSGAMPNQAQMRILKETELRKVKVLGSGAFQVYKGIWIPDGENVKIPVAIKVLREN 758
Db 687 PLGP-GEKGTVKHARILKPSDLRKIKPLGSGVGTGSKGFWPEGETVKIPVAIKTIQDS 745
QY 759 TSPKANKEILDEAYVMAGVSPVSVRLGLICLTSTVOLVTLMPYGLLDHVRNRRGLG 818
Db 746 SGRQTFTETDHLSSMGLDHPVIRLLGICPGTCLQVLTQLSSHGSLLEHSHRQKTSLD 805
QY 819 SODLLNMCIOAKGMSVLEDRVLVHRDLAARNVLKSPNKHVITDFGLARLLDIDETEVH 878
Db 806 PORLLNCVOIAGMYLEBHRVHKNLAARNILLKNQYQVQISDYGVADLLVPDDKKYV 865
QY 879 ADGKVPKIMWALESILRRFTHQSDVMSYGVTVWELMTFCAPYDGIIPAREIPDLLEKG 938
Db 866 YSEKTKPIKMALESILFRYTHQSDVMSYGVTVWEMWFGABYASVQVQVPSVLEKG 925
QY 939 ERLPQPPICIDVYIMVWKWIDSECRPRELVSFERSMARDPQRFVVIQNEIDLGPAS 998
Db 926 ERLSQAICIDVYIMVWKWIDENIRPTFKELASDFTRMARDPPRYLVIRMEG----- 980
QY 999 PLDSTFVRSLEDDMDGLVDAEYLVPOQGFPCDPAPGAGGMVHRHSSSTRSGGD 1058
Db 981 -----EUGMGFEFL-----RGSER---GL 997
QY 1059 LTLGLEPSEBEAPRSLAPSEGAGSDVFDGLDGMG---AAKGLQSLPTHDPSPQLQ----- 1110
Db 998 LEADLEDEEE-----GLGDRFATPSLQPSWSWTSQSINSYM 1036
QY 1111 -----RYSEDPVLPSETGYVAPLTCSPQ- EYVNO-----PDVRPQP 1150
Db 1037 VMTQLRYD-----FAVSOGGHIGYLPMSPPVDITRQLWYQSRSLSSVRTLPRSAFR 1090
QY 1151 SPREGPL--PAARPAGATLERAKTSLSPKNGVVKDVFAFGAVENPEYLTPOGGAPOPH 1208
Db 1091 SSREALECEDGAQACAGIFRVR-----FGSERGN-----POGG----- 1122
QY 1209 PPAPFAPFNLVYWDQPPERGAPPSTFKGTPTAENPE 1247
Db 1123 -----QORKLSTASSPSPFKTWADEEDE 1146

RESULT 7

Q9BIH9
ID Q9BIH9 PRELIMINARY; PRT; 1433 AA.
AC Q9BIH9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;

OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SU4;
RA Lycett G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
epidermal growth factor receptor."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ301655; CAC35008.1; --
DR HSP; P11362; 1FGK
DR InterPro; IPR000345; Cytochrome bind.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TYRKc; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS0011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor..
FT NON_TER 1
SQ SSQUENCE 1433 AA; 159585 MW; E3D9D8967724F07 CRC64;

Query Match 29.4%; Score 2000.5; DB 5; Length 1433;
Best Local Similarity 32.6%; Pred. No. 2.8e-142;
Matches 464; Conservative 191; Mismatches 383; Indels 385; Gaps 36;

QY 50 YGCGVQVQGNLEIYLTPTNASISFLQDIOEVQGYVLIHNRQVRQVPLRLIRVGRGTQLF- 108
Db 25 YTNCTVVDGNLEITWQNTIDNLNLFQIHREVTGYVLISLVLDLPQVILPRLQIIRGRTTK 84
QY 109 ----ENYALAVLDNDGDLNNTPTVTGASPGGLRELQRLSLTEILKGVLIQRNPOLCYQ 164
Db 85 LNKWEAYGLFV-----SFSHMTLELPALRDLILGSGVGFNNYLNCHM 128
QY 165 DTILWKDI-FHKNNQLALTIDNRSRACHPCSPCKSGSCWGESSEDCQSLTRTYVCAGG 223
Db 129 KSNWEEILLAPQTSQYTFNFSSPVRVCPCHPSCEVG-CWGEHANCQRFSLKNCSPQ 187
QY 224 CA--RCKGPLETDCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFES 281
Db 188 CSQGRCFGPKPRECCHLFCAGGCTGPTQSDCLACKNFYDDGVCKQECPPMQIYNPTNFW 247
QY 282 MNPEGRYTFGASCTVACPYNYLSTDVSGCTLVCPHNOEVTAEADTQRCCKSKPCARV 341
Db 248 ENPDPGKAYGATCVKRCP-EHLLKDNAGCAVRCKPKGMPQNSE-----CVPCKGVCPKT 301
QY 342 CYCLGMHLEHREVRVTSANIQFAGCKTIFGSLAFLPESFGDGPASNT-----APLQ 393
Db 302 CPGEIVH-----SDNIGNYKDCITIEGSLLEILDQSFQDGFQVYTFNPSFGPRYIKID 353
QY 394 PEOLQVFETLEEITGYLYISAWPDSLPDLSVFONLQVIRGRIIHNGAY-SLTQGLGISW 452
Db 354 PDELEVFSVKEITGFINTQAHHPNFTTLNFRNLEVGVGRQLKENLFSASVIVKTSLSK 413
QY 453 LGLRSRLRGSLGALIHNNHLCFVHTVPDQGLFRNPHOALLHTANRPEDECVGEGGLACH 512
Db 414 LELSLKRVNSGSIIVLENSDLCFVEDIDWSBKSSDEHVMQKRNATECHEEGMECS 473
QY 513 QLCARGHCWGPPTQCVNCSQFLRGQECVEECRVLOGLPREY-VNARHCLPCHPECPQN 571
Db 474 EQCSKAGCWGKPEQCLECKNVKYGKCLUDSK---SLPRLYSVDSKTCGDCHQECCKD-- 528

```
Qy 572 GSVTCFGEADQCACAHYKDPFCVACRP----- 601
Db 529 --FCYGNEDCGSCNWKDGRFCVACPTTKHAMNGTCINCHKTCVCGRGRDITIAPD 585
Qy 602 -----SGVKPDLSYMPIWKFPD-----BEG----- 621
Db 586 GCISCDKAIIGSDAKIERLKMDESCPDGYYSYDVLOBEGPLKQLSGKAVCRKCHPRCKK 645
Qy 622 -----ACOPCPINCT-----HSC 634
Db 646 CTGTFHFQFQCECTGYKKGQCEDECPQDPYANEETRICUPCHQECRGCHGLDDHHEC 705
Qy 635 VDL-----DD-----KGCPAEQ-----RASPLTSI 654
Db 706 RNLKLFEGDPYDNATTFVCVSNCPASHPHYKFPQEGAGKIGYCSADSMQSGRLRTEPQTQV 765
Qy 655 VSAVVGILLVVVLGVVFGI-----LKRQOKIRKYTMRRLLQETELVPLPTSPGAMPNOAQ 711
Db 766 KIVMGSVNALILLVCVFGIAFVFSRHKNKDDAVKMTMALAGCEDSEPLRPSNVGPNLT 825
Qy 712 MRILKETELRKVKVLSGAGFCVYKGIWIPDGENVKIPVALKVLRENTSPKANKEIILDEA 771
Db 826 LRITKEAIRGGVGLMGAFGRVFGVMWPGESVKIPVALKVLMEWSGSESSKEFELEA 885
Qy 772 YWAGVGSYVSRILIGICTVOLTQVLMPEYGCILLDHVRENRRGLSGQDILLNMCQIAK 831
Db 886 YIMASVEHPNLLKLLAVCMTSQMMLITQMLPLGCLLDYVRNNKDKIGSKALLNWSQIAR 945
Qy 932 GMSYLEVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKMAL 891
Db 946 GMAYLEERRLVHRDLAARNVLVPTSCVKITVFGKLKLLDPSDEYRAAGGKMPDKMAL 1005
Qy 892 ESILRRRPTHOSDVWSYGVTVWELMTGAKYDGI PAREIPDLLEKGERLPPOPICTIDV 951
Db 1006 ECIRHRVFTSKSDVWAFGITTWELTYGARDYENVPKDVPELIEIGHKUPQPDICSLDV 1065
Qy 952 YIMVVKCWMIDSECRPRELVSFBSFMRARDPQRFVVIQNEBDLGPASPLDSTFYRSLLD 1011
Db 1066 YCILLSCVLDADARPTFKQLAETFAEKARDPGYLM-----PGDKFELPSVTN 1116
Qy 1012 DDMGDLV-----DAEVLVPOQGFPCDDPAPGAGGM 1042
Db 1117 QDEKDLIRTLAPVAMAAAAAGASNVDPSTIAETDEYLQPKTRPSIMLPGPSA--- 1173
Qy 1043 VHRHRSSSTRSGGDLTLGLEPSEERAPRS-----PLAP---SEGAGSDVFDGDLGMG 1093
Db 1174 -----VEPS-DEMPKSLRYCKDPLKDDTDGKGKEV-----GVG 1207
Qy 1094 AAKGLQSLPTHDPSPLORYSDPTVPLPSETDGYVAPLTCSPQPEYVNPQDVRPQPPSPR 1153
Db 1208 GIR-----LNLPLDEDDYLP-TCOSQ---NQS----- 1231
Qy 1154 EGPLPAARPAGATLERAKTSLPGKNGVVKOVFAFGGAVENPEYL-----TPQGGAA 1204
Db 1232 -----TPG-----YMDLIGVPASVDNPEYLMGSTQAIAGLAQSGMG 1267
Qy 1205 PQPHPPPAFSPADNLYWDQDPPERGAPSTFKGTPTAENPE 1247
Db 1268 --PHTPP-----PPNTPNGMPTHQHSQ 1287

RESULT 8
ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
OX NCBI_TaxID=9606;
RN [1]
RX SEQUENCE FROM N.A.
MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
  autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).
RN [2]
RX SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
EMBL; AF177761; AAD56009.2; -
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FEC1BE347E2D030C CRC64;

Query Match 25.9%; Score 1762.5; DB 4; Length 419;
Best Local Similarity 84.7%; Pred. No. 4.7e-125;
Matches 337; Conservative 9; Mismatches 37; Indels 15; Gaps 4;

Qy 1 MELAALCRWGLLLALLPPGAASQTQCTENNFTVSFVLRVPKVSASHLE-----LYQGQCVV 56
Db 1 MELAALCRWGLLLALLPPGAASQTQCTGDMK-----LRLPASPTHLDMLRHLXQGQCVV 56
Qy 57 QGNLELYLPTNASLFLQDIOEVQGVYLIAHNQVQVPLQRLRIVRGTLQFEDNYALAV 116
Db 57 QGNLELYLPTNASLFLQDIOEVQGVYLIAHNQVQVPLQRLRIVRGTLQFEDNYALAV 116
Qy 117 LDNGDPLNNTTPTVGASPGRLQLRSLEILLKGGVLTORNPOLCYQDITLWKDIFHKN 176
Db 117 LDNGDPLNNTTPTVGASPGRLQLRSLEILLKGGVLTORNPOLCYQDITLWKDIFHKN 176
Qy 177 NQALALITDNRSRACHPCSPCKGSRGWESSEDCSLTRTVTCAGGCARCKGLPTDCC 236
Db 177 NQALALITDNRSRACHPCSPCKGSRGWESSEDCSLTRTVTCAGGCARCKGLPTDCC 236
Qy 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCV 296
Db 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCV 296
Qy 297 TACPYNVLSVDVSCSTLVCPLNHNOETAEADGTORCEKCKPCARVCYGLGMEHLREVAV 356
Db 297 TACPYNVLSVDVSCSTLVCPLNHNOETAEADGTORCEKCKPCARVCYGLGMEHLREVAV 356
Qy 357 TSANIOEFAGCKKIFGSLAPLPESFDGDPASNTAPLQP 394
Db 352 VVPLRWQPG--PAHPVLSFLRPSWDLVSFAFYSPLAP 387

RESULT 9
Q8R2X1 PRELIMINARY; PRT; 367 AA.
ID Q8R2X1
AC Q8R2X1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RX SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
EMBL; BC027080; AAH27080.1; -
KW Hypothetical protein.
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;
```



```
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
KW Tyrosine-protein kinase.
FT NON TER 1
SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725B1 CRC64;

Query Match 25.28; Score 1718; DB 15; Length 567;
Best Local Similarity 55.44; Pred. No. 1.7e-121;
Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY 578 GPEADQCACAHYKDPFCVAPCGSVKPDLSYPIWKFPDEEGACQPCINCTHSCVDL 637
DB 1 GP--DHCWKCAHFIDGCHVCACAGVLENDTL-VWKYADANAVCQLCHPCNTRGCKGP 57

QY 638 DRKCPAEQASPLTISVAVV-GILLVVLGVVGVGILIKRQOKIRKYMYRRLLOTEL 696
DB 58 GLEGCP---NGSKTPSTAAGVVGGLCLVVGVLGIGLYLRRR-HIVRKRTLRLLOTEL 113

QY 697 VEPLTPSGAMPNQAEILKTELKVKYLGSGAGFTVYKGIWIPDGENVKIPVAIKVLR 756
DB 114 VEPLTPSGEAPNQAEILKTELKVKYLGSGAGFTVYKGIWIPDGENVKIPVAIKELR 173

QY 757 ENTSPKANKELIDAYVMAGVGSYVSRLLGICLTSTVQLVTLMPYGCCLLDHVRENRR 816
DB 174 ENTSPKANKELIDAYVMSVDNPRVCELLGICLTSTVQLITQLMPYGCCLLDVIREHKN 233

QY 817 LGSQDLLNWCQIAKGSYLEDLVRLVHRDLAARNVLKSNHVKITDFGLARLLIDETE 876
DB 234 IGSQDLLNWCQIAKGNVLEERLVRDLAARNVLKTPQHKVITDFGLAKLLGADEKE 293

QY 877 YHAGGKVPKWALESILRRFTHQSDVMSYGVTVWELMTFGAKPDGIPAREIPDLLE 936
DB 294 YHAGGKVPKWALESILHRIYTHQSDVMSYGVTVWELMTFGSKPDGIPASEISVLE 353

QY 937 KGERLPOPPICTIDVYIMVKWMDISECRPRFRELVSERFMRARDPQRFVVIQ-NEDLG 995
DB 354 KGERLPOPPICTIDVYIMVKWMDADSRKRELIAEFSKWARDPPRVLVQGDREMH 413

QY 996 PASPLDSTFRSLDDMDGLVDAEYLVPOQGFPCPDAPAGGVMVHRRSSSTRSG 1055
DB 414 LPSPTDSKFRYLTMEEDMEDIVDAEYLVPHQGF-----NSPST---454

QY 1056 GGDLTGLSEEEAPSP-----APSEGAGSDVFDGLGMAKGLQSLPTHDPSPLO 1110
DB 455 -----SRTPLLSSLSATSNNSATNCID-----RNGQGHVPREDSFVQ 491

QY 1111 RYSEDPTVPLPSET--DGYVAPLTCSQPQYVNVQPDVVRPQPPSPREGPLPAARPAGATLE 1168
DB 492 RYSDPTGNFLESIDGFL-----PAPEYVNO--LMPKKPS-----526

QY 1169 RAKTLPKGNKGVKDVF-----AFGGAVENPEYL 1197
DB 527 ----TAMVQNGIYNNISLTAISKLPMDSRQYNSHSTAVDNPEYL 566

RESULT 12
Q8WYV0 PRELIMINARY; PRT; 412 AA.
AC Q8WYV0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 44.7 kDa protein.
GN PF3659.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
```

```
RA Wan D.F., Gu J.R.;
RT "Novel human cDNA clones with function of inhibiting cancer cell growth."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF318349; AAL55856.1; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00669; pkinase; 1.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD00001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00118; EF HAND; UNKNOWN 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW Hypothetical protein.
SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match 24.9%; Score 1697.5; DB 4; Length 412;
Best Local Similarity 80.5%; Pred. No. 3.9e-120;
Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

QY 889 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPDGIPAREIPDLLEKGERLPQPPICT 948
DB 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPDGIPAREIPDLLEKGERLPQPPICT 60

QY 949 IDVYIMVKWMDISECRPRFRELVSERFMRARDPQRFVVIQNEDELGASPLDSTFVRS 1008
DB 61 IDVYIMVKWMDISECRPRFRELVSERFMRARDPQRFVVIQNEDELGASPLDSTFVRS 120

QY 1009 LEDDDMDGLVDAEYLVPOQGFPCPDAPAGGVMVHRRSSSTRSGGDLTLGLLEPSEE 1068
DB 121 LEDDDMDGLVDAEYLVPOQGFPCPDAPAGGVMVHRRSSSTRSGGDLTLGLLEPSEE 180

QY 1069 EAPRSLAPSEAGSDVFDGLGMAKGLQSLPTHDPSPLOQRYSDPTVPLPSETDGYV 1128
DB 181 EAPRSLAPSEAGSDVFDGLGMAKGLQSLPTHDPSPLOQRYSDPTVPLPSETDGYV 240

QY 1129 APLTCSQPQYVNVQPDVVRPQPPSPREGPLPAARPAGATLERAKTLPKGNKGVKDVFAFG 1188
DB 241 APLTCSQPQYVNVQPDVVRPQPPSPREGPLPAARPAGATLERAKTLPKGNKGVKDVFAFG 300

QY 1189 GAVENPEYLTPOGGAAPQ-----HPPPA---ESPAPDNL 1220
DB 301 GAVENPEYLTPOGGAALPTLLPSAQPSITGTRTHQSGGLHPAPSKGLRQRTQST 360

QY 1221 YWVD-QDPPER-----GAPPSFTFGTPTAEN 1245
DB 361 WWTQCEPEGVRRSPDVSSGREGLTSAKIKRWEGPPTTSRGTCARN 410

RESULT 13
Q64895 PRELIMINARY; PRT; 962 AA.
AC Q64895;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Gag, v-erb-A, v-erb-B protein.
GN GAG, V-ERB-A, V-ERB-B.
OS Avian erythroblastosis virus.
OC Viruses; Retroviridae; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxId=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9026603; PubMed=1969616;
RT Bruskina A., Jackson J., Bishop J.M., McCarley D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the transforming potential of the oncogene v-erb-B."
RT Oncogene 5:15-24 (1990).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; X52209; CAA36459.1; -.
```

```
DR EMBL; X52211; CAA36459.1; JOINED.
DR HSP; P10828; 2NLL.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Sterhorm_receptor.
DR InterPro; IPR001245; Tyr_kinase.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD000001; Euk_pkinase; 1.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOLI; 1.
DR SMART; SM00219; TyfK; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW Transcription regulation; Transferase; Tyrosine-protein kinase;
KW Zinc-finger
SQ SEQUENCE 962 AA; 108320 MW; 3CSAED791E4E95CE CRC64;

Query Match 24.3%; Score 1653.5; DB 15; Length 962;
Best Local Similarity 51.7%; Pred. No. 2.9e-116;
Matches 358; Conservative 73; Mismatches 142; Indels 119; Gaps 18;

QY 541 VESCRVLQGLPRE-VYNAR-HCLP-----CHPEQC 568
DB 354 IERCOESYLLAFEHYNYRHNIPHFWSKLLMKVADLRMIGAYHASRFLHMKVECPTELS 413
QY 569 PONGSVTCFGEADQCACAHYKDPFCVACRCPGVKPDLSYPIKFPDEEGACQPCPI 628
DB 414 PQE-----VGP---DHCWKCAHFIDGPHCVKACPAVLGENDTL-VWKYADANAVCQLCHP 465
QY 629 NCTHSCVDLDDKCPAEQRASPLTSIVSAVV-GILLVVLGVVFGILIKRQOKIRKYTM 687
DB 466 NCTRGCKGPGLEGCP---NGSKTPSIAAGVVGGLCLVVGLGIGLYLRR-HIVRKRTL 521
QY 688 RLLQETELVEPLTPSGAMPNQAMRILKETELRKVKVLGSGAGFTVYKGIWIIPDGENVK 747
DB 522 RLLQRELVEPLTPSGEAPNQAHRLILKETEFKVKVLGFGAGFTVYKGLWIPEGEKVT 581
QY 748 IPVAIKVLRNTSPKANKEILDEAYVMAGVSPVSRLLGICLTSTVOLVTLQMPYGCCL 807
DB 582 IPVAIKELRATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVOLVTLQMPYGCCL 641
QY 808 DHVRENRLGSLDNLNWCQIAKMSYLEDRVLRVHRDLAARNVLKSPNHVKITDFGLA 867
DB 642 DYIREHKDNGISQYLLNWCQIAKGMNLYLBERHVRDLAARNVLKTPQHVKITDFGLA 701
QY 868 RLLDIDETEYHAGGKVPKWMALLESILRRFTHQSDVWSYGVTVWELMTFGAKPYDGP 927
DB 702 KQLGADEKEYHAEGBGVPIKWMALLESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGP 761
QY 928 AREIPDLLEGERLPQPPICITIDVYIMVKCMWIDSECRPRFRELSEFSRMARDPQRFV 987
DB 762 ASEISSVLEGERLPQPPICITIDVYIMVKCMWSDASRPKRFRELIAEFSKWARDPPRYL 821
QY 988 VIO-NEDLGASPDLSTFYRSLLDEDDMDGLVDAEYLVPOQGFCCPDAPGAGGMVHR 1046
DB 822 VIQDERMHLPSPTDSKFYRTLNEEDMEDIVDAEYLVPHQGF-----866
QY 1047 HRSSSTRSGGDLTLGLEPSEEBEAPRSPAPSGAGSDVFDGLGMAKAGLSLTHDP 1106
DB 867 -NSPST-----SRTPLLSLSLTSN-----NSATKCIDRNGCH-- 898
QY 1107 SPLQRVSEDTVPPLPSETDGYVAPLTCSPQEVYVQPDVPRPQPPSPREGPLPAARAGAT 1166
DB 899 -----PVREDGFL-----PAPEYVNO--LMPKKPSTAMVQNIYINISLT 936
```

```
QY 1167 -LERAKTSPGKNGVGVKDVFAFGGAVENPEYL 1197
DB 937 AISKLPMDSRYN-----SHSTAVDNPEYL 961

RESULT 14
Q85468 PRELIMINARY; PRT; 545 AA.
ID Q85468;
AC Q85468;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Avian Erythroblastosis virus (Ts34) v-erbB gene.
OS Avian erythroblastosis virus.
OC Viruses; Retrovirdae; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=88217326; PubMed=2897102;
RP Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
RT virus mutants that are temperature sensitive for transformation.";
RL Oncogene Res. 1:285-278(1987).
DR EMBL; X06943; CAA30024.1; -.
DR HSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyfK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 545 AA; 60899 MW; 140DCBCCA0F8AF4 CRC64;
```

```
Query Match 24.1%; Score 1645; DB 15; Length 545;
Best Local Similarity 54.9%; Pred. No. 5.5e-116;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

QY 578 GPEADQCACAHYKDPFCVACRCPGVKPDLSYMPIKFPDEEGACQPCPINTHSCVDL 637
DB 1 GP--DHCWKCAHFIDGPHCVKACPAVLGENDTL-VWKYADANAVCQLCHPNCTRGCKGP 57
QY 638 DKGCPAEQRASPLTSIVSAVV-GILLVVLGVVFGILIKRQOKIRKYTMERLLQETEL 696
DB 58 GLEGCP---NGSKTPSIAAGVVGGLCLVVGLGIGLYLRR-HIVRKETLRLLQEREL 113
QY 697 VEPLTPSGAMPNQAMRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVL 756
DB 114 VEPLTPSGEAPNQAHRLILKETEFKVKVLGFGAGFTVYKGLWIPEGEKVTIPVAIKEL 173
QY 757 ENTPSKANKEILDEAYVMAGVSPVSRLLGICLTSTVOLVTLQMPYGCCLDHVRENRR 816
DB 174 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVOLVTLQMPYGCCLDVIREHKD 233
QY 817 LGSQDLNWCQIAKMSYLEDRVLRVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 876
DB 234 IGSQYLLNWCQIAKGMNLYLBERHVRDLAARNVLKTPQDVKITDFGLAQQLGADEKE 293
QY 877 YHADGKVPKWMALLESILRRFTHQSDVWSYGVTVWELMTFGAKPYDGPAREIPDLLE 936
DB 294 YHAEGBGVPIKWMALLESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGPASEISSVLE 353
QY 937 KGERLPQPPICITIDVYIMVKCMWIDSECRPRFRELSEFSRMARDPQRFVVIQ-NEDLG 995
DB 354 KGERLPQPPICITIDVYIMVKCMWSDASRPKRFRELIAEFSKWARDPPRYLVIQGDERMH 413
QY 996 PASPLDSTFYRSLLDEDDMDGLVDAEYLVPOQGFCCPDAPGAGGMVHRHRSSTRSG 1055
DB 414 LPSPTDSKFYRTLNEEDMEDIVDAEYLVPHQGF-----NSPST---- 454
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 37.9774 Seconds
(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-5-25-14
Perfect score: 6812
Sequence: 1 MELALCRWGLLLALLPPGA.....TFKGTPTAENPYLGLDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6700	98.4	1255	21	Human heregulin 2
2	6700	98.4	1255	22	Human tyrosine kin
3	6700	98.4	1255	22	HER2 transgene pla
4	6700	98.4	1255	23	Human HER2 (erbB2)
5	6694	98.3	1255	17	HER-2/neu protein.
6	6694	98.3	1255	20	Human HER-2/neu on
7	6694	98.3	1255	21	Human HER-2/neu pr
8	6694	98.3	1255	21	Amino acid sequenc
9	6694	98.3	1255	22	Human HER-2/neu pr
10	6694	98.3	1255	22	HER2/neu amino aci

11	6694	98.3	1255	23	AAE24067	Human Her-2 protei
12	6694	98.3	1255	23	AAE20479	Human Her-2/neu pr
13	6694	98.3	1255	23	AAE20479	Human Her-2/neu pr
14	6694	98.3	1255	23	AAU77114	Human Her-2/neu po
15	6651	97.6	1233	14	AAE39568	Sequence of c-erbB
16	6572	96.5	1223	23	AAU98923	Human breast cance
17	6377	93.6	1200	21	AAE21208	Human HER-2/neu pr
18	5886.5	86.4	1256	21	AAE21199	Rat HER-2/neu prot
19	5886.5	86.4	1256	23	AAE21144	Rat Her-2/neu onco
20	5861.5	86.0	1256	21	AAE21206	Mouse Her-2/neu pr
21	5861.5	86.0	1256	22	AAE20860	Amino acid sequenc
22	5861.5	86.0	1256	23	AAE21151	Mouse Her-2/neu on
23	4780	70.2	919	21	AAE21203	Human HER-2/neu fu
24	4780	70.2	919	23	AAE21148	Mouse Her-2/neu ex
25	4011.5	58.9	920	23	AAE21152	Mouse Her-2/neu ex
26	4011.5	58.9	926	23	AAE21153	Mouse Her-2/neu ex
27	3664	53.8	712	21	AAE21204	Human HER-2/neu fu
28	3664	53.8	712	23	AAE21149	Her-2/neu extracel
29	3518	51.6	782	18	AAE19764	Her2-GM-CSF immuno
30	3516	51.6	553	21	AAE21200	Extracellular HER-
31	3516	51.6	553	23	AAE21145	Human ErbB2 oncopr
32	3478	51.1	545	22	AAE60408	Human ErbB2 extrac
33	3478	51.1	545	22	AAE61593	Human ErbB2 extrac
34	3413	50.1	551	21	AAE44993	DC8cFv-erbB2EC fu
35	3310	48.6	524	11	AAE08222	Extracellular port
36	3132	46.0	1210	21	AAE19259	Amino acid sequenc
37	3132	46.0	1210	21	AAE19259	Human EGF receptor
38	3132	46.0	1210	23	AAE23019	Human Her-1 protei
39	3132	46.0	1210	23	AAE50768	Human epidermal gr
40	3130	45.9	1210	22	AAE68420	Amino acid sequenc
41	3091	45.4	1210	23	ABP51768	Human epidermal gr
42	3084	45.3	583	23	AAE20483	Human protein for
43	3084	45.3	587	23	AAE20481	Human protein for
44	3083	45.3	589	23	AAE20484	Human protein for
45	3083	45.3	600	23	AAE20482	Human protein for

ALIGNMENTS

RESULT 1
AAE24067
ID AAY92620 standard; Protein; 1255 AA.
XX AC AAY92620;
XX AC AAY92620;
DT 10-AUG-2000 (first entry)
XX Human heregulin 2 (Her2).
DE Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
XX Homo sapiens.
XX Key Location/Qualifiers
FH Domain 1..173
FT /label= N-terminal
FT /note= "mature polypeptide"
FT 5..25
FT Region /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 59..73
FT Region /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 103..117
FT Region /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 149..163
FT Region /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 174..323
FT Domain

FT	Region	/label= Cysteine_rich_domain 210..324	CC	is used for inducing immune responses against weakly immunogenic cell-associated peptide antigens (PA) such as those associated with
FT	Region	/label= insertion region	CC	cancers (self-proteins), e.g. human prostate specific membrane antigen
FT	Region	/note= "suitable for foreign epitope insertion"	CC	(PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
FT	Region	250..264	CC	The method comprises effecting simultaneous presentation by antigen
FT	Region	/label= insertion region	CC	producing cells (APCs) of the animals immune system of: (1) at least 1
FT	Region	/note= "suitable for foreign epitope insertion"	CC	CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
FT	Region	324..483	CC	B-cell group derived from the cell-associated PA; and (2) at least 1
FT	Region	325..339	CC	first T helper cell group which is foreign to the animal. Analogues of
FT	Region	/label= insertion region	CC	human PSM, human Her2 and human/murine FGF8b comprising a substantial
FT	Region	/note= "suitable for foreign epitope insertion"	CC	part of all known and predicted CTL and B-cell epitopes of the respective
FT	Region	369..383	CC	PA and including at least one foreign T helper epitope are also claimed.
FT	Region	/label= insertion region	CC	The method is used to treat prostate, prostate/breast or breast cancer
FT	Region	/note= "suitable for foreign epitope insertion"	CC	when the PA is human PSM, FGF8b and Her2, respectively.
FT	Region	465..479	XX	
FT	Region	/label= insertion region	SQ	Sequence 1255 AA;
FT	Region	/note= "suitable for foreign epitope insertion"		
FT	Region	484..623		Query Match 98.4%; Score 6700; DB 21; Length 1255;
FT	Region	579..593		Best Local Similarity 98.4%; Pred. No. 0;
FT	Region	/label= Cysteine_rich_domain		Matches 1239; Conservative 4; Mismatches 8; Indels 8; Gaps 2;
FT	Region	/label= insertion region		
FT	Region	/note= "suitable for foreign epitope insertion"		
FT	Region	624..654		
FT	Region	/label= Transmembrane_domain		
FT	Region	632..652		
FT	Region	/label= insertion region		
FT	Region	/note= "suitable for foreign epitope insertion"		
FT	Region	653..667		
FT	Region	/label= insertion region		
FT	Region	/note= "suitable for foreign epitope insertion"		
FT	Region	655..1010		
FT	Region	/label= Tyrosine_kinase_domain		
FT	Region	661..675		
FT	Region	/label= insertion region		
FT	Region	/note= "suitable for foreign epitope insertion"		
FT	Region	695..709		
FT	Region	/label= insertion region		
FT	Region	/note= "suitable for foreign epitope insertion"		
FT	Region	710..730		
FT	Region	/label= insertion region		
FT	Region	/note= "suitable for foreign epitope insertion"		
FT	Region	1011..1235		
FT	Region	/label= C-terminal_domain		
XX	WO200020027-A2.			
XX	13-APR-2000.			
XX	05-OCT-1999;	99WO-DK00525.		
XX	05-OCT-1998;	98DK-0001261.		
XX	20-OCT-1998;	98US-0105011.		
XX	(NEBI-) M & E BIOTECH AS.			
XX	Steinaa L, Mouritseen S, Nielsen KG, Haaning J, Leach D, Dalum I;			
XX	Gautam A, Birk P, Karlsson G;			
XX	WPI; 2000-349917/30.			
XX	N-PSDB; AAA09455.			
XX	Inducing immune responses to weakly immunogenic, tumor associated			
XX	peptide antigens for the treatment of breast and prostate cancer			
XX	Claim 62; Page 193-198; 220pp; English.			
XX	This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of			
XX	Her2 can be used in the claimed method as an autovaccine to induce a CTL			
XX	response. Subdominant CTL epitopes, antibody binding regions and			
XX	cysteine residues involved in disulfide bonds are preserved in the			
XX	immunogenized forms. Regions suitable for the insertion of foreign T			
XX	helper epitopes were identified (see features table). The method			

QY 777 VGSPPYRLIGICLTSTVQLVTOLMPYGCILLDHVRNRLGSLGSDLLNWCMTAKMSYL 836
 Db 777 VGSPPYRLIGICLTSTVQLVTOLMPYGCILLDHVRNRLGSLGSDLLNWCMTAKMSYL 836
 QY 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALLESILR 896
 Db 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALLESILR 896
 QY 897 RRETHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYIMV 956
 Db 897 RRETHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYIMV 956
 QY 957 KCMWIDSECRPRELVSFESRMARDOPRVVIONEDLGPASPLDSTFYRSILLEDDMDG 1016
 Db 957 KCMWIDSECRPRELVSFESRMARDOPRVVIONEDLGPASPLDSTFYRSILLEDDMDG 1016
 QY 1017 LVDAAEYLVPOQGFCDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEAAPRPLA 1076
 Db 1017 LVDAAEYLVPOQGFCDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEAAPRPLA 1076
 QY 1077 PSEGAGSDVFDGLGMAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQ 1136
 Db 1077 PSEGAGSDVFDGLGMAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQ 1136
 QY 1137 PEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKGVWVDVAFGAVENPEY 1196
 Db 1137 PEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKGVWVDVAFGAVENPEY 1196
 QY 1197 LTPQGGAAQPQHPPPAFSPAFDNLVYWDQDPPRGAPPSTFKTPTAENPEYLGLDVPV 1255
 Db 1197 LTPQGGAAQPQHPPPAFSPAFDNLVYWDQDPPRGAPPSTFKTPTAENPEYLGLDVPV 1255

RESULT 2

AAE12130
 ID AAE12130 standard; Protein; 1255 AA.

XX AAE12130;

DT 18-DEC-2001 (first entry)

XX Human tyrosine kinase-type receptor, HER-2.

XX Therapeutic compound; major histocompatibility complex; vaccine;
 KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
 KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
 XX antigen presenting cell; human; tyrosine kinase-type receptor.

OS Homo sapiens.

XX Key Location/Qualifiers
 FH Region 774..782
 FT /note= "Antigenic epitope"

XX WO200168677-A2.

XX 20-SEP-2001.

XX 16-MAR-2001; 2001WO-US40328.

XX 16-MAR-2000; 2000US-0527487.

XX (GENZ) GENZYME CORP.

XX Nicolette CA;

XX WPI; 2001-616284/71.

DR N-PSDB; AAD19731.

XX Novel synthetic therapeutic compound for inducing immune response and
 PT for use in adoptive immunotherapy, has enhanced binding to major
 PT histocompatibility molecules and enhanced immunoregulatory properties

XX Claim 4; Page 63-67; 69pp; English.

XX The invention relates to synthetic therapeutic compounds (antigenic
 CC peptides) with enhanced binding to major histocompatibility complex
 CC (MHC) molecules and enhanced immunoregulatory properties relative
 CC to their natural counterparts. Compounds of the invention are useful
 CC for inducing an immune response in a subject and for use in adoptive
 CC immunotherapy. They are useful as components of anti-cancer vaccines
 CC and to expand immune effector cells that are specific for cancers
 CC characterised by expression of the breast cancer antigen, HER-2.
 CC Polynucleotides that encode peptides of the invention are useful as
 CC hybridisation probes and as primers for the detection of genes of gene
 CC transcripts that are expressed in antigen presenting cells (APCs), to
 CC confirm transduction of polynucleotides into host cells. The present
 CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
 CC of the invention are designed based on the HER-2 antigenic peptide
 CC (774-782).

XX Sequence 1255 AA;

Query Match 98.4%; Score 6700; DB 22; Length 1255;

Best Local Similarity 98.4%; Pred. No. 0;

Matches 1239; Conservative 4; Mismatches 8; Indels 8; Gaps 2;

QY 1 MELAALCRWGLLLALLPPGAASTQVCTFNNFTVSWLRVVKVSASHLE----LYQGCQVY 56

Db 1 MELAALCRWGLLLALLPPGAASTQVCTGTDNK---LRLPASPETHLDMLRHLVYQGCQV 56

QY 57 QGNLELTYPNTNASLFLQDIOEVGYVLIHAHQVQVPLQRLRVRGTLFEDNYALAV 116

Db 57 QGNLELTYPNTNASLFLQDIOEVGYVLIHAHQVQVPLQRLRVRGTLFEDNYALAV 116

QY 117 LDNGDPLNNTTPTVTCASPGGLRELQRLSLTELKGGVLIQRNPOLCYODTILWKDIFHKN 176

Db 117 LDNGDPLNNTTPTVTCASPGGLRELQRLSLTELKGGVLIQRNPOLCYODTILWKDIFHKN 176

QY 177 NQLALTLIDTNRSRACHPCSMKGRCSWGSESDCQSLTRTVCAAGGCARCKGLPTDCC 236

Db 177 NQLALTLIDTNRSRACHPCSMKGRCSWGSESDCQSLTRTVCAAGGCARCKGLPTDCC 236

QY 237 HEQCAAGCTGFKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTGASCV 296

Db 237 HEQCAAGCTGFKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTGASCV 296

QY 297 TACPNYLTSTVGSCTLVCPHLNQSVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAV 356

Db 297 TACPNYLTSTVGSCTLVCPHLNQSVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAV 356

QY 357 TSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFTLEETITGYLYISAMP 416

Db 357 TSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFTLEETITGYLYISAMP 416

QY 417 DSLPDLSPVQMLQVIRGRILHNGAYSLTQGLGISWLGRLSRLRELGLALIHNNTHLCF 476

Db 417 DSLPDLSPVQMLQVIRGRILHNGAYSLTQGLGISWLGRLSRLRELGLALIHNNTHLCF 476

QY 477 VHTVPWDOLFNRPHQALLHTANRPEDECVGEGELACHQLCARGHCWGPGTCQVNCQSOFILR 536

Db 477 VHTVPWDOLFNRPHQALLHTANRPEDECVGEGELACHQLCARGHCWGPGTCQVNCQSOFILR 536

QY 537 GQECVVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPDFC 596

Db 537 GQECVVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPDFC 596

QY 597 VARCPGSKPDLISYMPIWKFPEDEGACQCPINCTHSCVDLDDKCPAEQASPLTSTVS 656

Db 597 VARCPGSKPDLISYMPIWKFPEDEGACQCPINCTHSCVDLDDKCPAEQASPLTSTVS 656

QY 657 AVVGILLVVVLGVVFGIILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILK 716

Db 657 AVVGILLVVVLGVVFGIILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILK 716

```

QY 717 ETELKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAG 776
DB |||||
QY 717 ETELKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAG 776
DB |||||
QY 777 VGSPPYVSRLLGICLTSTVQLVTQMPYGCCLLDHVRENRLGSGDILLNWCQIAKMSYL 836
DB |||||
QY 777 VGSPPYVSRLLGICLTSTVQLVTQMPYGCCLLDHVRENRLGSGDILLNWCQIAKMSYL 836
DB |||||
QY 837 EDVRLVHRDLAARNVLKSNHVKITDFGLARLLDIDETEHADGGKVPKKNWALSILR 896
DB |||||
QY 837 EDVRLVHRDLAARNVLKSNHVKITDFGLARLLDIDETEHADGGKVPKKNWALSILR 896
DB |||||
QY 897 RRFTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGBRLPQPPICTIDVYMIW 956
DB |||||
QY 897 RRFTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGBRLPQPPICTIDVYMIW 956
DB |||||
QY 957 KCMWIDSECHPRFRELVSERWARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMGD 1016
DB |||||
QY 957 KCMWIDSECHPRFRELVSERWARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMGD 1016
DB |||||
QY 1017 LVDAEEYLVPOQGFPCDPAPGAGGVMHRRSSSTRSGGDLTLGLEPSEEAAPRPLA 1076
DB |||||
QY 1017 LVDAEEYLVPOQGFPCDPAPGAGGVMHRRSSSTRSGGDLTLGLEPSEEAAPRPLA 1076
DB |||||
QY 1077 PSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLOQYSEDPTVLPSETDGYVAPLTCSPQ 1136
DB |||||
QY 1077 PSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLOQYSEDPTVLPSETDGYVAPLTCSPQ 1136
DB |||||
QY 1137 PEYVQPDVPRPQPPSPREGPLPAARAGATLERAKTILSPGKNGVVDVFAFGAVENPEY 1196
DB |||||
QY 1137 PEYVQPDVPRPQPPSPREGPLPAARAGATLERAKTILSPGKNGVVDVFAFGAVENPEY 1196
DB |||||
QY 1197 LTPQGGAAPOPHPPAFSPAFDNLXYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
DB |||||
QY 1197 LTPQGGAAPOPHPPAFSPAFDNLXYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
DB |||||

RESULT 3
AAB60167
ID AAB60167 standard; Protein; 1255 AA.
XX AC AAB60167;
XX AC AAB60167;
XX AC AAB60167;
DT 03-APR-2001 (first entry)
DE HER2 transgene plasmid construct encoded protein.
XX KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
XX KW antibody.
XX OS Homo sapiens.
XX OS Synthetic.
XX FN WO200100244-A2.
XX PD 04-JAN-2001.
XX PF 23-JUN-2000; 2000WO-US17229.
XX PR 25-JUN-1999; 99US-0141316.
XX PR 16-MAR-2000; 2000US-0189844.
XX PA (GETH ) GENENTECH INC.
XX PI Erickson S, Schwall R;
XX DR WPI; 2001-061962/07.
XX DR N-PSDB; AAF24297.
XX PT Treating tumors, particularly breast cancers, which overexpress an ErbB
XX PT receptor and does not respond to an anti-ErbB antibody, comprises
XX PT conjugating the antibody to a maytansinoid -
XX

```

Example 3; Fig 4; 92pp; English.

The present invention provides a method of treating cancer by administering a conjugate of anti-ErbB antibody with a maytansinoid. In particular, the antibody is directed against ErbB2 (also known as HER2 and p185neu). The method is particularly useful in the treatment of breast, ovarian, stomach, endometrial, salivary gland, lung, kidney, colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.

Sequence 1255 AA;

Query Match 98.4%; Score 6700; DB 22; Length 1255;

Best Local Similarity 98.4%; Pred. No. 0;

Matches 1239; Conservative 4; Mismatches 8; Indels 8; Gaps 2;

```

QY 1 MELAALCRWGLLLALLPPGAASSTVCTFNNFTVSFWLRVPKVSASHLE---LYQGCQV 56
DB |||||
QY 1 MELAALCRWGLLLALLPPGAASSTVCTGDMK----LRLPASPTHDLMLRHLVQGCQV 56
DB |||||
QY 57 QGNLELTPTNASLSFLQDIOEVGYVLIANQVRQVPLQRLRIVRGTLQEDNVALAV 116
DB |||||
QY 57 QGNLELTPTNASLSFLQDIOEVGYVLIANQVRQVPLQRLRIVRGTLQEDNVALAV 116
DB |||||
QY 117 LDNGDPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHN 176
DB |||||
QY 117 LDNGDPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHN 176
DB |||||
QY 177 NQALATLDTNRSRACHPCSMKSCRCWGESSEDCQSLTRTVCAAGCARCKPLPTDCC 236
DB |||||
QY 177 NQALATLDTNRSRACHPCSMKSCRCWGESSEDCQSLTRTVCAAGCARCKPLPTDCC 236
DB |||||
QY 237 HQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGYTFGASCV 296
DB |||||
QY 237 HQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGYTFGASCV 296
DB |||||
QY 297 TACPYNYLSTDVGSCTLVCPHLNQVTAEDGTORCEKSKPCARVCYGLGMEHLREVRV 356
DB |||||
QY 297 TACPYNYLSTDVGSCTLVCPHLNQVTAEDGTORCEKSKPCARVCYGLGMEHLREVRV 356
DB |||||
QY 357 TSANIOEFAGCKKIFGSLAFIPESFDGDPASNTAPLOPELOVFTLEITGYLISAWP 416
DB |||||
QY 357 TSANIOEFAGCKKIFGSLAFIPESFDGDPASNTAPLOPELOVFTLEITGYLISAWP 416
DB |||||
QY 417 DSLPDLVSFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLGSLALHNNTHLCF 476
DB |||||
QY 417 DSLPDLVSFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLGSLALHNNTHLCF 476
DB |||||
QY 477 VHTVPWDQLFRNPHQALLHTANRPEDECVGEGLAGCHQLCARGHCWGPPTQCVNCSQFLR 536
DB |||||
QY 477 VHTVPWDQLFRNPHQALLHTANRPEDECVGEGLAGCHQLCARGHCWGPPTQCVNCSQFLR 536
DB |||||
QY 537 GOECVEECRVLOGLPREYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPC 596
DB |||||
QY 537 GOECVEECRVLOGLPREYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPC 596
DB |||||
QY 597 VARCPGKVPDLSPYMPIWKFPDEEGACOPCPINCTHSCVDLDDKGCAPAEQASPLTSIVS 656
DB |||||
QY 597 VARCPGKVPDLSPYMPIWKFPDEEGACOPCPINCTHSCVDLDDKGCAPAEQASPLTSIVS 656
DB |||||
QY 657 AVVGILLVVVLGVVFGILIKRRQQKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILK 716
DB |||||
QY 657 AVVGILLVVVLGVVFGILIKRRQQKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILK 716
DB |||||
QY 717 ETELKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAG 776
DB |||||
QY 717 ETELKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAG 776
DB |||||
QY 777 VGSPPYVSRLLGICLTSTVQLVTQMPYGCCLLDHVRENRLGSGDILLNWCQIAKMSYL 836
DB |||||
QY 777 VGSPPYVSRLLGICLTSTVQLVTQMPYGCCLLDHVRENRLGSGDILLNWCQIAKMSYL 836
DB |||||
QY 837 EDVRLVHRDLAARNVLKSNHVKITDFGLARLLDIDETEHADGGKVPKKNWALSILR 896
DB |||||

```

Db 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVKPIKMWALESIILR 896
QY 897 RRETHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGBRLPQPPICTIDVYMIWV 956
Db 897 RRETHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGBRLPQPPICTIDVYMIWV 956
QY 957 KCMWIDSECHPRELVSFESRWARDPQRFVVIQNEEDLGPASPLDSTFYRSLLLEDDMGD 1016
Db 957 KCMWIDSECHPRELVSFESRWARDPQRFVVIQNEEDLGPASPLDSTFYRSLLLEDDMGD 1016
QY 1017 LVDABEYLVPOQGFCDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEAAPRSPLA 1076
Db 1017 LVDABEYLVPOQGFCDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEAAPRSPLA 1076
QY 1077 PSEGAGSDVFDGLMGAAKGLQSLPHTDPSPLQRYSEDPVLPSETDGYVAPLTCSPQ 1136
Db 1077 PSEGAGSDVFDGLMGAAKGLQSLPHTDPSPLQRYSEDPVLPSETDGYVAPLTCSPQ 1136
QY 1137 PEYVNPQDVRPQPPSPREGPLPAARAGATLEAKTILSPKNGVVKDVFAGGAVENPEY 1196
Db 1137 PEYVNPQDVRPQPPSPREGPLPAARAGATLEAKTILSPKNGVVKDVFAGGAVENPEY 1196
QY 1197 LTPQGAAPQPHPPPAFSPAFDNLVYWDQDPPPERGAPPSTFKTPTAENPEYLGLDVPV 1255
Db 1197 LTPQGAAPQPHPPPAFSPAFDNLVYWDQDPPPERGAPPSTFKTPTAENPEYLGLDVPV 1255
RESULT 4
AAU74545
ID AAU74545 standard; Protein; 1255 AA.
XX
AC AAU74545;
XX
DT 23-APR-2002 (first entry)
XX
DE Human HER2 (ErbB2) polypeptide.
XX
KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;
KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
KW glial disorder; astrocytal disorder; hypothalamic disorder;
KW glandular disorder; macrophagal disorder; epithelial disorder;
KW stromal disorder; blastocoelec disorder; inflammatory disorder;
KW angiogenic disorder; immunological disorder.
XX
OS Homo sapiens.
XX
PN US2002001587-A1.
XX
PD 03-JAN-2002.
XX
PF 16-MAR-2001; 2001US-0811123.
XX
PR 16-MAR-2000; 2000US-189844P.
PR 05-OCT-2000; 2000US-238327P.
XX
PA (ERIC/) ERICKSON S.
PA (SCHW/) SCHWALL R.
PA (SLIW/) SLIWKOWSKI M.
XX
PI Erickson S, Schwall R, Sliwowski M;
XX WPI; 2002-163686/21.
DR N-PSDB; ABK14058.
XX
PT Treating tumour characterised by overexpression of epidermal growth
PT factor receptor, ErbB or cancer in mammal, comprises administering
PT anti-ErbB antibody-maytansinoid conjugate to the mammal -
XX
PS Example 3; Fig 7; 93pp; English.
XX
CC The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor
CC receptor (ErbB) and does not respond or responds poorly, to treatment
CC with an anti-ErbB antibody, comprising administering to the mammal an
CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
CC prostate and bladder, preferably breast cancer. The breast cancer is a
CC metastatic breast cancer or an aggressive form of metastatic breast
CC cancer which overexpresses ErbB2. The method is also useful for treating
CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
CC epithelial, stromal, blastocoelec, inflammatory, angiogenic and
CC immunological disorders. This sequence represents the human HER2 (ErbB2)
CC polypeptide of the invention.
XX
SQ Sequence 1255 AA;
Query Match 98.4%; Score 6700; DB 23; Length 1255;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 1239; Conservative 4; Mismatches 8; Indels 8; Gaps 2;
QY 1 MELAALCRWGLLLALLPPGAASTQVCTFNNFTVSFMLRVPKVSASHLE----LYQGCQVW 56
Db 1 MELAALCRWGLLLALLPPGAASTQVCTGDMK---LRLPASPEHLDMLRHLYQGCQVW 56
QY 57 QCNLELTYPNNAISFLQDIOEVQGYVLIQNVQVPLQRLIRVRGTQLFEDNVALAV 116
Db 57 QCNLELTYPNNAISFLQDIOEVQGYVLIQNVQVPLQRLIRVRGTQLFEDNVALAV 116
QY 117 LDNGDPLANNTPVTGASPGGLRELOLRSITELTKGGVLIQNPOLCYQDTILKWDIFHN 176
Db 117 LDNGDPLANNTPVTGASPGGLRELOLRSITELTKGGVLIQNPOLCYQDTILKWDIFHN 176
QY 177 NQLALTLDITRSRACHPCSPMKGSRGWGESSEDCQSLTRTVACGGACARCKGPIPTDCC 236
Db 177 NQLALTLDITRSRACHPCSPMKGSRGWGESSEDCQSLTRTVACGGACARCKGPIPTDCC 236
QY 237 HQCAAGCTGPKHSDCLACLHPNHSGICELHCPALVTYNTDTFESMPNPEGRYTGASCV 296
Db 237 HQCAAGCTGPKHSDCLACLHPNHSGICELHCPALVTYNTDTFESMPNPEGRYTGASCV 296
QY 297 TACPYNILSTDVGSCTLCPLHNOVTAEDGTORCEKSKPCARVCYGLGMEHLREVAV 356
Db 297 TACPYNILSTDVGSCTLCPLHNOVTAEDGTORCEKSKPCARVCYGLGMEHLREVAV 356
QY 357 TSANIOEFAGCKKIFGSLAFLPESPDGPASNTAPLOPELOVVFETLEITGYLISAMP 416
Db 357 TSANIOEFAGCKKIFGSLAFLPESPDGPASNTAPLOPELOVVFETLEITGYLISAMP 416
QY 417 DSLPDLVSFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLRELGSLALIHNNTHLCF 476
Db 417 DSLPDLVSFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLRELGSLALIHNNTHLCF 476
QY 477 VHTVPMDQLFRNPQALLHTANRPEDECVGEGELACHOLCARGHCWGPPTQCVCNSQFLR 536
Db 477 VHTVPMDQLFRNPQALLHTANRPEDECVGEGELACHOLCARGHCWGPPTQCVCNSQFLR 536
QY 537 GOECVBECKRVLOGLPREVYNARHCLPCHPECOQNGSVTCFGEADQCVACAHYKDPFPC 596
Db 537 GOECVBECKRVLOGLPREVYNARHCLPCHPECOQNGSVTCFGEADQCVACAHYKDPFPC 596
QY 597 VARCPGKVPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIVS 656
Db 597 VARCPGKVPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIVS 656
QY 657 AVVIGILLVVVLGVVFGIILKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILK 716
Db 657 AVVIGILLVVVLGVVFGIILKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILK 716
QY 717 ETELAKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAG 776
Db 717 ETELAKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAG 776
QY 777 VGSYPVSRLLIGLICLTSTVQLVTQMLPYPGCLLDHVRENRGRGLSGQDLNWCQIAKMSYL 836

```

Db 777 VGSPPVSRLLGICLTSTVQLVTQMPYGCCLDHVRENRGRSGSODLLNMCQIAKMSYL 836
QY 837 EDVRLVHRDLAARNLVKSNHVKITDGLARLLDIDETEHADGGKVPKKNWALSILR 896
Db 837 EDVRLVHRDLAARNLVKSNHVKITDGLARLLDIDETEHADGGKVPKKNWALSILR 896
QY 897 RRETHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGRRLPQPPICTIDVYMIW 956
Db 897 RRETHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGRRLPQPPICTIDVYMIW 956
QY 957 KCWIDSECRPRFRELVSFMRWDPQRFVVIQNEEDLGPASPLDSTFYRSLLDDDDMGD 1016
Db 957 KCWIDSECRPRFRELVSFMRWDPQRFVVIQNEEDLGPASPLDSTFYRSLLDDDDMGD 1016
QY 1017 LVDAEYLYPQOGFFCPDPAAGAGVHHRRSSSTRSGGDLTLGLESEEEAPRSPLA 1076
Db 1017 LVDAEYLYPQOGFFCPDPAAGAGVHHRRSSSTRSGGDLTLGLESEEEAPRSPLA 1076
QY 1077 PSEGAGSDVFDGLGMAKGLQSLPHDPSPLQRYSEDPVLPSETDGYVAPLTCSPQ 1136
Db 1077 PSEGAGSDVFDGLGMAKGLQSLPHDPSPLQRYSEDPVLPSETDGYVAPLTCSPQ 1136
QY 1137 PEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVFAFGAVENPEY 1196
Db 1137 PEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVFAFGAVENPEY 1196
QY 1197 LTPQGGAAQPHPPPAFSPAFDNLYYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
Db 1197 LTPQGGAAQPHPPPAFSPAFDNLYYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

```

RESULT 5

AAW01111 standard; Protein; 1255 AA.

AC AAW01111;

DT 01-JAN-1997 (first entry)

DE HER-2/neu protein.

XX HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
 KW breast cancer; ovary cancer; colon cancer; lung cancer;
 KW prostate cancer; immunisation; tumour; vaccine; vector.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Domain 676..1255

FT /label= Intracellular domain

FT /note= "claimed domain, useful for immunisation"

XX W09630514-A1.

XX 03-OCT-1996.

XX 28-MAR-1996; 96WO-US01689.

XX 31-MAR-1995; 95US-041417.

XX (UNIW) UNIV WASHINGTON.

XX Cheever MA, Disis ML;

XX WPI; 1996-455361/45.

XX N-PSDB; AAT40739.

XX DNA encoding HER-2-neu poly.peptide(s) - used for prevention or
 PT treatment of malignancies with which the HER-2/neu oncogene is
 PT associated

XX Claim 2; Page 56-61; 71pp; English.

XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
 CC the product of the HER-2/neu oncogene (see also AAT40739). The
 CC protein is over-expressed in various cancers, including breast,
 CC ovarian, colon, lung and prostate. The intracellular domain of the
 CC protein can be used to immunise an animal against a malignancy with
 CC which the oncogene is associated. The polypeptide can be produced
 CC in transformed host cells for use in immunisation. Alternatively,
 CC animal cells are transfected in vivo or ex vivo with a viral vector
 CC that directs expression of the polypeptide.

XX SQ Sequence 1255 AA;

Query Match 98.3%; Score 6694; DB 17; Length 1255;

Best Local Similarity 98.3%; Pred. No. 0;

Matches 1237; Conservative 5; Mismatches 9; Indels 8; Gaps 2;

QY 1 MELAALCRWGLLLALLPPGAASTVCTFNNFTVSFMLRVPKVSASHLE----LYQSCQVV 56

Db 1 MELAALCRWGLLLALLPPGAASTVCTGTDK---LRLPASPTHLDMLRLHYQGCQVV 56

QY 57 QGNLELTYPNTNASLSFLQDIOEVQYVLIANQVRQVPLQRLRIVRGTLFEDNYALAV 116

Db 57 QGNLELTYPNTNASLSFLQDIOEVQYVLIANQVRQVPLQRLRIVRGTLFEDNYALAV 116

QY 117 LONGDPLNNTTPTVTCASPGGLREQLRSITELKGVLIQRPOLCYQDTILWKDIFHN 176

Db 117 LONGDPLNNTTPTVTCASPGGLREQLRSITELKGVLIQRPOLCYQDTILWKDIFHN 176

QY 177 NQALTLTIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRVCAGGCARCKGLPTDCC 236

Db 177 NQALTLTIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRVCAGGCARCKGLPTDCC 236

QY 237 HEQCAAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296

Db 237 HEQCAAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296

QY 297 TACPNYLYSTDVGSCTLVCPHNVQVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRV 356

Db 297 TACPNYLYSTDVGSCTLVCPHNVQVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRV 356

QY 357 TSANIQEFAGCKKIFGSLAFLPESFDGPASNTAPLQPEQLQVFETLEITGYLYISAMP 416

Db 357 TSANIQEFAGCKKIFGSLAFLPESFDGPASNTAPLQPEQLQVFETLEITGYLYISAMP 416

QY 417 DSLPDLVSFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLSGSLALIHNTLHLCF 476

Db 417 DSLPDLVSFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLSGSLALIHNTLHLCF 476

QY 477 VHTVPWDQLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWPGPTQCVCNCSQIFUR 536

Db 477 VHTVPWDQLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWPGPTQCVCNCSQIFUR 536

QY 537 GOECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVCACAHYKDPDPPFC 596

Db 537 GOECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVCACAHYKDPDPPFC 596

QY 597 VARCPGKVPDLISYMPIWKFDPDEGACQPCPINCTHSCVDLDDKCGPABQASPLTSIVS 656

Db 597 VARCPGKVPDLISYMPIWKFDPDEGACQPCPINCTHSCVDLDDKCGPABQASPLTSIVS 656

QY 657 AVVGILLVVVLGVVFGILIKRQOKIRKYTWBRLAQETELVEPLTPSGAMPNQAQMRILK 716

Db 657 AVVGILLVVVLGVVFGILIKRQOKIRKYTWBRLAQETELVEPLTPSGAMPNQAQMRILK 716

QY 717 ETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAG 776

Db 717 ETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAG 776

QY 777 VGSPVSRLLGICLTSTVQLVTQMPYGCCLDHVRENRGRSGSODLLNMCQIAKMSYL 836

Db 777 VGSPVSRLLGICLTSTVQLVTQMPYGCCLDHVRENRGRSGSODLLNMCQIAKMSYL 836

QY 837 EDVRLVHRDLAARNVLKSPNHVKITDPLGLARLLDIDETEHADGGKVPKIMWALSILR 896
DB 837 EDVRLVHRDLAARNVLKSPNHVKITDPLGLARLLDIDETEHADGGKVPKIMWALSILR 896
QY 897 RRTTHOSDWSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGRLLPQPPICITIDVYMIW 956
DB 897 RRTTHOSDWSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGRLLPQPPICITIDVYMIW 956
QY 957 KCMWIDSECRPRFRELVSFSESRWARDPQRFVVIQNEEDLGPASPLDSTFYRSLLDDDDMGD 1016
DB 957 KCMWIDSECRPRFRELVSFSESRWARDPQRFVVIQNEEDLGPASPLDSTFYRSLLDDDDMGD 1016
QY 1017 LVDAEYLYPQOQFFCPDPAPGAGGVHHRHRSSTRSGGDLTLGLEPSEEBAPRSPLA 1076
DB 1017 LVDAEYLYPQOQFFCPDPAPGAGGVHHRHRSSTRSGGDLTLGLEPSEEBAPRSPLA 1076
QY 1077 PSEGAGSDVFDGLGMAAGLQSLPHDPSPLQRYSEDTVPPLPSETDGYVAPLTCSPQ 1136
DB 1077 PSEGAGSDVFDGLGMAAGLQSLPHDPSPLQRYSEDTVPPLPSETDGYVAPLTCSPQ 1136
QY 1137 PEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPKNGVVKDVFAFGAVENPEY 1196
DB 1137 PEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPKNGVVKDVFAFGAVENPEY 1196
QY 1197 LTPQGAAPQPHPPAPFAPFNDLYYWDQDPPPERGAPPSTFKTPTAENPEYLGLDVPV 1255
DB 1197 LTPQGAAPQPHPPAPFAPFNDLYYWDQDPPPERGAPPSTFKTPTAENPEYLGLDVPV 1255

RESULT 6
AAW92406
ID AAW92406 standard; Protein; 1255 AA.
XX
AC AAW92406;
XX
DT 21-APR-1999 (first entry)
XX
DE Human HER-2/neu oncogene protein.
XX
KW HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
KW malignancy; treatment; tumour.
XX
OS Homo sapiens.
XX
FH Key
FT Region 676..1255
FT Location/Qualifiers
FT /note= "region which elicits immune response"
XX
PN US5869445-A.
XX
XX 09-FEB-1999.
XX
XX 01-APR-1996; 96US-0625101.
XX
XX 01-APR-1996; 96US-0625101.
PR 17-MAR-1993; 93US-0033644.
PR 12-AUG-1993; 93US-0106112.
PR 31-MAR-1995; 95US-0414417.
XX
XX (UNIW) UNIV WASHINGTON.
XX
XX Cheever MA, Disis ML;
XX
XX WPI; 1999-152835/13.
DR N-PSDB; AAX01912.
XX
XX Use of HER-2/neu polypeptides - for eliciting an immune response to
PT an HER-2/neu associated malignancy, particularly for treating or
PT preventing tumours
XX
XX Claim 3; Column 31-38; 26pp; English.
PS
XX This sequence represents the human HER-2/neu oncogene protein. A fragment
CC

CC of this protein is used in a method for eliciting or enhancing an immune
CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
CC B cells to produce an immune response to the HER-2/neu protein. The
CC method can be used for immunisation against a malignancy in which the
CC HER-2/neu oncogene is associated and in the treatment of an existing
CC tumour, or to prevent tumour occurrence or reoccurrence.
XX
SQ Sequence 1255 AA;
Query Match 98.3%; Score 6694; DB 20; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1237; Conservative 5; Mismatches 9; Indels 8; Gaps 2;
QY 1 MELAALCRWGLLLALLPPGAASTVCTFNNFVSFWLRVPKVSASHLE----LYQGCQVV 56
DB 1 MELAALCRWGLLLALLPPGAASTVCTGTDK-----LRLPASPEHLMDLRLHYQGCQVV 56
QY 57 QGNLELTYPNASTSFLQDIOEVQGYVLIANQVROPVLPQRLRIVRGTLFEDNYALAV 116
DB 57 QGNLELTYPNASTSFLQDIOEVQGYVLIANQVROPVLPQRLRIVRGTLFEDNYALAV 116
QY 117 LONGDPLNNTTPTVTCASPGGLRELQRLSLEILKGVLIQRPOLCYQDTILWKDIFHKN 176
DB 117 LONGDPLNNTTPTVTCASPGGLRELQRLSLEILKGVLIQRPOLCYQDTILWKDIFHKN 176
QY 177 NOLALTLDITNRSRACHPCSPCKGSRGWGSEDCQSLTRTVCAAGCARCKGPLTDCC 236
DB 177 NOLALTLDITNRSRACHPCSPCKGSRGWGSEDCQSLTRTVCAAGCARCKGPLTDCC 236
QY 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
DB 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
QY 297 TACPYNILSTVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVAV 356
DB 297 TACPYNILSTVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVAV 356
QY 357 TSANIQEFAGCKKIFGSLAFPESEFDGPASNTAPLQPEQLQVVFETLEITGYLISAMP 416
DB 357 TSANIQEFAGCKKIFGSLAFPESEFDGPASNTAPLQPEQLQVVFETLEITGYLISAMP 416
QY 417 DSLPDLSPVQNLQVIRGRILHNGAYSILQGLGISWLGRLSRLRELSGLALIHNTLHLCF 476
DB 417 DSLPDLSPVQNLQVIRGRILHNGAYSILQGLGISWLGRLSRLRELSGLALIHNTLHLCF 476
QY 477 VHTVWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPPTQCVNCSQFLR 536
DB 477 VHTVWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPPTQCVNCSQFLR 536
QY 537 GOECVBECEVRLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPC 596
DB 537 GOECVBECEVRLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPC 596
QY 597 VARPSPGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCAPQASPLTSIYS 656
DB 597 VARPSPGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCAPQASPLTSIYS 656
QY 657 AVVGILLVVVLGVVFEGLIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILK 716
DB 657 AVVGILLVVVLGVVFEGLIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILK 716
QY 717 ETELKRVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAG 776
DB 717 ETELKRVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAG 776
QY 777 VGSPPVSRLLGLTSTVQLVTLQMPYGLCLDHHVRENRGLSGDOLLNWCQIAKMSYL 836
DB 777 VGSPPVSRLLGLTSTVQLVTLQMPYGLCLDHHVRENRGLSGDOLLNWCQIAKMSYL 836
QY 837 EDVRLVHRDLAARNVLKSPNHVKITDPLGLARLLDIDETEHADGGKVPKIMWALSILR 896
DB 837 EDVRLVHRDLAARNVLKSPNHVKITDPLGLARLLDIDETEHADGGKVPKIMWALSILR 896

QY 897 RRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYMW 956
Db 897 RRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYMW 956
QY 957 KCMWIDSECHPRRELVSERMRARDPQRFVVIQNEEDLGPASPLDSTFYRSLLDDMDGD 1016
Db 957 KCMWIDSECHPRRELVSERMRARDPQRFVVIQNEEDLGPASPLDSTFYRSLLDDMDGD 1016
QY 1017 LVDAEYLVPOQGFCDPAPGAGGVHHRHRSSTRSGGDLTLGLEPSEERAPSPPLA 1076
Db 1017 LVDAEYLVPOQGFCDPAPGAGGVHHRHRSSTRSGGDLTLGLEPSEERAPSPPLA 1076
QY 1077 PSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDTVPLPSETDGYVAPLTCSPQ 1136
Db 1077 PSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDTVPLPSETDGYVAPLTCSPQ 1136
QY 1137 PEYVQPDVVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEY 1196
Db 1137 PEYVQPDVVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEY 1196
QY 1197 LTFQGAAPOPHPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKTPTAENPEYLGLDVPV 1255
Db 1197 LTFQGAAPOPHPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKTPTAENPEYLGLDVPV 1255

RESULT 7

AAB21198
ID AAB21198 standard; protein; 1255 AA.

AC AAB21198;

XX 12-JAN-2001 (first entry)

XX Human HER-2/neu protein.

XX Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
KW colon cancer.

XX Homo sapiens.

XX WO200044899-A1.

XX 03-AUG-2000.

XX 28-JAN-2000; 2000WO-US02164.

XX 29-JAN-1999; 99US-0117976.

XX (CORI-) CORIXA CORP.

XX (SMIK) SMITHKLINE BEECHAM.

XX Cheever MA, Gheyssen D;

XX WPI; 2000-505976/45.

XX N-PSDB; AAB89736.

XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT useful for vaccinating against breast, ovarian, colon, lung and
PT prostate cancers -

XX Claim 52; Fig 7; 128pp; English.

XX The present sequence is the human HER-2/neu protein. It is a member of
CC the tyrosine kinase family of receptor-like glycoproteins and shows
CC homology to the epidermal growth factor receptor (EGFR). It probably
CC plays a part in cell growth and/or differentiation. The HER-2/neu
CC gene is an oncogene. An HER-2/neu fusion protein comprising a
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC domain may be used to treat or prevent cancer by eliciting or
CC enhancing an immune response to the HER-2/neu protein. It may be used
CC to treat malignancies such as breast, ovarian, colon, lung and
CC prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.
XX
SQ Sequence 1255 AA;

Query Match 98.3%; Score 6694; DB 21; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1237; Conservative 5; Mismatches 9; Indels 8; Gaps 2;

QY 1 MELAALCRWGLLALLPPGAASCTVCTFNFTVSEFWLRVPKVSASHLE---LYQCQCVV 56
Db 1 MELAALCRWGLLALLPPGAASCTVCTGTDMK----LRUPASPETHDMLRHLYYQCQCVV 56
QY 57 QGNLEITYLPTNASLSFLQDIOEVQGYVLI AHNQVRQVPLQRLRIVRGTFQEDNYALAV 116
Db 57 QGNLEITYLPTNASLSFLQDIOEVQGYVLI AHNQVRQVPLQRLRIVRGTFQEDNYALAV 116
QY 117 LONGDPLNNTTPTVGTASPGGLRELQRLSRLTEILKGGVLIQRNPOLCYQDTILWKDIFHKN 176
Db 117 LONGDPLNNTTPTVGTASPGGLRELQRLSRLTEILKGGVLIQRNPOLCYQDTILWKDIFHKN 176
QY 177 NQALATLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCC 236
Db 177 NQALATLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCC 236
QY 237 HEQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
Db 237 HEQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
QY 297 TACPNYILSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAV 356
Db 297 TACPNYILSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAV 356
QY 357 TSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFFETLEITGVLYISAMP 416
Db 357 TSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFFETLEITGVLYISAMP 416
QY 417 DSLPDLVSFQNLQVIRGRILHNGAYSRLTQGLGISWLGLRSRLRELSGLALIHNTLHLCF 476
Db 417 DSLPDLVSFQNLQVIRGRILHNGAYSRLTQGLGISWLGLRSRLRELSGLALIHNTLHLCF 476
QY 477 VHTVPWQDLFRNPHQALLHTANRPEDECVEGLACHQLCARGHCWGPGTQCNCVSQFLR 536
Db 477 VHTVPWQDLFRNPHQALLHTANRPEDECVEGLACHQLCARGHCWGPGTQCNCVSQFLR 536
QY 537 GOECVEECRVLQGLPREYVNAHCLPCHPECOQNGSVTCFCGEADQCACAHYKDPFPC 596
Db 537 GOECVEECRVLQGLPREYVNAHCLPCHPECOQNGSVTCFCGEADQCACAHYKDPFPC 596
QY 597 VARCPGKVPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCAPQASPLTSIVS 656
Db 597 VARCPGKVPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCAPQASPLTSIVS 656
QY 657 AVVGILLVVVLGVVFGILIKRQOKIRKYTMRLRLQETELVEPLTPSGAMPNOAQRILK 716
Db 657 AVVGILLVVVLGVVFGILIKRQOKIRKYTMRLRLQETELVEPLTPSGAMPNOAQRILK 716
QY 717 ETELKRVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPPKANKEIIDEAYVMAG 776
Db 717 ETELKRVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPPKANKEIIDEAYVMAG 776
QY 777 VGSPYVSRLIGLICLTSTVQLVTLQMPYGCCLLDHVRENRGLSGQDLLNWCMIAGMSYL 836
Db 777 VGSPYVSRLIGLICLTSTVQLVTLQMPYGCCLLDHVRENRGLSGQDLLNWCMIAGMSYL 836
QY 837 EDVRLVHRDLAARNLVKSPNHVKITDPLGLARLLDIDETEHADGGKVPKIMWALESILR 896
Db 837 EDVRLVHRDLAARNLVKSPNHVKITDPLGLARLLDIDETEHADGGKVPKIMWALESILR 896
QY 897 RSEFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYMW 956
Db 897 RSEFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYMW 956
QY 957 KCWMIDSECHPRRELVSERMRARDPQRFVVIQNEEDLGPASPLDSTFYRSLLDDMDGD 1016

Db 957 KCWMDISECRPRELVSFSESRWARDQREVVVQNEIDLGPASPLDSTFYKSLLEDMDGD 1016
QY 1017 LVDAEEYLVPOQGFPCDPAPGAGGVMVHHRHRSSTRSGGDLTLGLEPSEEEAPRPLA 1076
Db 1017 LVDAEEYLVPOQGFPCDPAPGAGGVMVHHRHRSSTRSGGDLTLGLEPSEEEAPRPLA 1076
QY 1077 PSEGAGSDVPDGLGMAAGLQSLPTHDPSPLOQRYSEDTVPPLPSETDGYVAPLTCSPQ 1136
Db 1077 PSEGAGSDVPDGLGMAAGLQSLPTHDPSPLOQRYSEDTVPPLPSETDGYVAPLTCSPQ 1136
QY 1137 PEYVNPDPVRPQPPSPREGPLPAARPAAGATLERAKTILSPCKNGVWVDVFAFGAVENPEY 1196
Db 1137 PEYVNPDPVRPQPPSPREGPLPAARPAAGATLERAKTILSPCKNGVWVDVFAFGAVENPEY 1196
QY 1197 LTPQGGAAPQHPPPAFSPAFDNLYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
Db 1197 LTPQGGAAPQHPPPAFSPAFDNLYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 8
AAY84780
ID AAY84780 standard; Protein; 1255 AA.
AC AAY84780;
XX 08-AUG-2000 (first entry)
DT Amino acid sequence of the SPLICE erbB-2 receptor protein.
DE SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;
KW tumor cell proliferation; tissue degeneration; arthropathy;
KW bone resorption; inflammatory disease; degenerative disorder;
KW wound healing.
XX Homo sapiens.
OS
XX
PN W0200020579-A1.
XX
PD 13-APR-2000.
XX
XX 01-OCT-1999; 99WO-CA00912.
XX
PR 02-OCT-1998; 98US-0165192.
XX
XX (UYMC-) UNIV MCMASTER.
XX
XX Muller WJ, Siegel PM;
XX
XX WPI; 2000-303768/26.
DR N-PSDB; AAA14812.
XX
XX Nucleic acid encoding an erbB 2 receptor protein designated SPLICE
PT erbB-2, inhibitors of the protein are useful for treatment of cancer -
XX
PS Claim 3; Fig 2; 60pp; English.
XX
XX The present sequence represents a SPLICE erbB-2 receptor protein. The
CC protein has an in-frame deletion of 16 amino acids, 2 of which are
CC conserved cysteine residues, compared to the unspliced protein. The
CC erbB-2 polynucleotide is used to construct probes for detecting
CC disorders of cell transformation such as cancer. Antibodies to the
CC protein may be used to detect SPLICE erbB-2 in a sample. Agents
CC (e.g. antisense oligonucleotides) which inhibit the expression of
CC SPLICE erbB-2 are useful for reducing tumor cell proliferation and
CC treating cancer. Substances which stimulate SPLICE erbB-2 are useful
CC for treating conditions involving damaged cells including conditions
CC in which degeneration of tissue occurs, such as arthropathy, bone
CC resorption, inflammatory diseases, degenerative disorders of the
CC central nervous system and wound healing.
XX
XX Sequence 1255 AA;

Query Match 98.3%; Score 6694; DB 21; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1237; Conservative 5; Mismatches 9; Indels 8; Gaps 2;
QY 1 MELAALCRWGLLALLPPGAASSTQVCTFNNFVSFWLRVPKVSASHLE---LYQCQCVV 56
Db 1 MELAALCRWGLLALLPPGAASSTQVCTGTDK-----LRUPASPETHLDMRLHLYQCQCVV 56
QY 57 QGNLELYLPTNASLSFLQDIOEVQGYVLIHAHQVRQVPLQRLIRVRGTQLFEDNYALAV 116
Db 57 QGNLELYLPTNASLSFLQDIOEVQGYVLIHAHQVRQVPLQRLIRVRGTQLFEDNYALAV 116
QY 117 LONGDPLNNTTPTVGTASPGGLRELQRLSLTEILKGVLIQRNPOLCYQDTILWKDIFHN 176
Db 117 LONGDPLNNTTPTVGTASPGGLRELQRLSLTEILKGVLIQRNPOLCYQDTILWKDIFHN 176
QY 177 NQALTLIDTNSRACHPCSPCKGSRGWSESDCOSLTRVTCAGGCARCKGLPTDCC 236
Db 177 NQALTLIDTNSRACHPCSPCKGSRGWSESDCOSLTRVTCAGGCARCKGLPTDCC 236
QY 237 HEQCAAGCTGFKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYFCASCV 296
Db 237 HEQCAAGCTGFKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYFCASCV 296
QY 297 TACPYNLYSTDVSGSCTLVCPHNOEVTABDGTQRCCKSKPCARVCYGLGMEHLREVRV 356
Db 297 TACPYNLYSTDVSGSCTLVCPHNOEVTABDGTQRCCKSKPCARVCYGLGMEHLREVRV 356
QY 357 TSANTQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPELOVFFETLEITGYLYISAMP 416
Db 357 TSANTQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPELOVFFETLEITGYLYISAMP 416
QY 417 DSLPDLVSFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLSGSLALHNNHLCF 476
Db 417 DSLPDLVSFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLSGSLALHNNHLCF 476
QY 477 VHTVPWDQLFENPHOALLHTANRPDECEVGEGLACHOLCARGHCGWPGPTQCVNCSQFLR 536
Db 477 VHTVPWDQLFENPHOALLHTANRPDECEVGEGLACHOLCARGHCGWPGPTQCVNCSQFLR 536
QY 537 GQECVEECRVLOGLPREYVNAHRCILPCHPECPQNGSVTCFGEADQCVACAHYKDPPEC 596
Db 537 GQECVEECRVLOGLPREYVNAHRCILPCHPECPQNGSVTCFGEADQCVACAHYKDPPEC 596
QY 597 VARCPGKVPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGPAPQASPLTSIVS 656
Db 597 VARCPGKVPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGPAPQASPLTSIVS 656
QY 657 AVVGILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILK 716
Db 657 AVVGILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILK 716
QY 717 ETELRKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAG 776
Db 717 ETELRKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAG 776
QY 777 VGSPVSVRLGLCLTSTVOLVTQLMPYGLLDHVRNRCRLGSDLLNWCMIAGMSYL 836
Db 777 VGSPVSVRLGLCLTSTVOLVTQLMPYGLLDHVRNRCRLGSDLLNWCMIAGMSYL 836
QY 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKWMALESILR 896
Db 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKWMALESILR 896
QY 897 RRPHTQSDVWSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPDPPCTIDVYIMV 956
Db 897 RRPHTQSDVWSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPDPPCTIDVYIMV 956
QY 957 KCWMDISECRPRELVSFSESRWARDQREVVVQNEIDLGPASPLDSTFYKSLLEDMDGD 1016
Db 957 KCWMDISECRPRELVSFSESRWARDQREVVVQNEIDLGPASPLDSTFYKSLLEDMDGD 1016
QY 1017 LVDAEEYLVPOQGFPCDPAPGAGGVMVHHRHRSSTRSGGDLTLGLEPSEEEAPRPLA 1076

Db 1017 LVDAEEYLVPOQGFCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEARPSPLA 1076
QY 1077 PSEGAGSDVFDGDLGMAAGLQSLPHDPSPLQRYSEDTVPPLPSETDGYVAPLTCSPQ 1136
Db 1077 PSEGAGSDVFDGDLGMAAGLQSLPHDPSPLQRYSEDTVPPLPSETDGYVAPLTCSPQ 1136
QY 1137 PEYVNPQDVVRPQPPSPREGPLPAARPAAGATLERAKTILSPKNGVVKDVFAFGGAVENPEY 1196
Db 1137 PEYVNPQDVVRPQPPSPREGPLPAARPAAGATLERAKTILSPKNGVVKDVFAFGGAVENPEY 1196
QY 1197 LTPQGAAPQHPHPPAFSPAFDNLYYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255
Db 1197 LTPQGAAPQHPHPPAFSPAFDNLYYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255

RESULT 9
AAB85458
ID AAB85458 standard; Protein: 1255 AA.
XX AC AAB85458;
XX AC AAB85458;
XX AC AAB85458;
DT 25-SEP-2001 (first entry)
XX Human HER-2/neu protein.
XX Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
XX oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX Homo sapiens.
XX WO200153463-A2.
XX 26-JUL-2001.
XX 19-JAN-2001; 2001WO-US01850.
XX 21-JAN-2000; 2000US-0177545.
XX (CORI-) CORIXA CORP.
XX Cheever MA, Hand-Zimmermann S;
XX WPI; 2001-476112/51.
XX N-PSDB; AAH23392.
XX New antigen-presenting cells, useful as vaccines for eliciting or
XX enhancing an immune response to HER-2/neu protein, particularly useful
XX for treating or preventing cancer, e.g. breast cancer -
XX Claim 2; Page 41-46; 49pp; English.
XX The invention provides an isolated antigen-presenting cell, which
XX expresses at least an immunogenic portion of a polypeptide that produces
XX an immune response to HER-2/neu protein. The antigen-presenting cells are
XX useful as vaccines for eliciting or enhancing an immune response to
XX HER-2/neu protein, particularly in treating or preventing malignancies in
XX which the HER-2/neu oncogene is associated. Specifically, these are
XX useful for treating or preventing cancer, e.g. breast cancer, ovarian,
XX colon, lung or prostate cancers. The present sequence represents
XX the human HER-2/neu protein (also known as p185 or c-erbB2).
XX Sequence 1255 AA;

Query Match 98.3%; Score 6694; DB 22; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1237; Conservative 5; Mismatches 9; Indels 8; Gaps 2;
QY 1 MELAALCRWGLLALLPPGAASTQVCTFNNFTVSWFLRVPKVSASHLE---LYQGCQVV 56
Db 1 MELAALCRWGLLALLPPGAASTQVCTGDMK---LRLPASPETHDMLRHLVYQGCQVV 56
QY 57 QGNLELTYPNTASLSFLQDIQEVQGYVLIAHNQVRQVPLQLRLIRVGTQLFEDNYALAV 116

Db 57 QGNLELTYPNTASLSFLQDIQEVQGYVLIAHNQVRQVPLQLRLIRVGTQLFEDNYALAV 116
QY 117 LONGDPLNNNTTVPVTGASPGGLRELQLRSUTEILKGGVLIQRPOLCYQDTILWKDIFHKN 176
Db 117 LONGDPLNNNTTVPVTGASPGGLRELQLRSUTEILKGGVLIQRPOLCYQDTILWKDIFHKN 176
QY 177 NOLAULTIDTNSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAGGCARCKGPLETDC 236
Db 177 NOLAULTIDTNSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAGGCARCKGPLETDC 236
QY 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
Db 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
QY 297 TACPYNLYSTDVSGSCTLVCPHNEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVAV 356
Db 297 TACPYNLYSTDVSGSCTLVCPHNEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVAV 356
QY 357 TSANTIQEPAGCKKIFGSLAPLPESPDGDPASNTAPLOPELOQVFEITGLYIISAMP 416
Db 357 TSANTIQEPAGCKKIFGSLAPLPESPDGDPASNTAPLOPELOQVFEITGLYIISAMP 416
QY 417 DSLPDLVSFQNLQVIRGRILHNGAYSILTQGLGISWGLRSRLSGSLALIHNNTHLCF 476
Db 417 DSLPDLVSFQNLQVIRGRILHNGAYSILTQGLGISWGLRSRLSGSLALIHNNTHLCF 476
QY 477 VHTVPMQDLFRPHQALLHTANRPEDECYGEGLACHOLCARGHCWGPCTQCNCQFLR 536
Db 477 VHTVPMQDLFRPHQALLHTANRPEDECYGEGLACHOLCARGHCWGPCTQCNCQFLR 536
QY 537 GQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPC 596
Db 537 GQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPC 596
QY 597 VARCPGSKPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCAPQASPLTSIYS 656
Db 597 VARCPGSKPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCAPQASPLTSIYS 656
QY 657 AVVGILLVVVLGVVFGILIKRQOKIRKYTMERLLQETELVEPLTPSGAMPNQAMRILK 716
Db 657 AVVGILLVVVLGVVFGILIKRQOKIRKYTMERLLQETELVEPLTPSGAMPNQAMRILK 716
QY 717 ETELKRVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPKANKEIIDEAAYMAG 776
Db 717 ETELKRVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPKANKEIIDEAAYMAG 776
QY 777 VGSPVSRLLGICLTSTVOLVTQMLPYGCLLDHVRNRLGSLQDLNWCMIAGMSYL 836
Db 777 VGSPVSRLLGICLTSTVOLVTQMLPYGCLLDHVRNRLGSLQDLNWCMIAGMSYL 836
QY 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKMALESILR 896
Db 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKMALESILR 896
QY 897 RFTTHQSDVWSVGVVWELMTGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMV 956
Db 897 RFTTHQSDVWSVGVVWELMTGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMV 956
QY 957 KCWIMIDSECRPRFRELVSFSEFSRMARDPQRFVVIQNEGLGSPASPLDSTFYRSLLEDMDGD 1016
Db 957 KCWIMIDSECRPRFRELVSFSEFSRMARDPQRFVVIQNEGLGSPASPLDSTFYRSLLEDMDGD 1016
QY 1017 LVDAEYLVPOQGFCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEARPSPLA 1076
Db 1017 LVDAEYLVPOQGFCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEARPSPLA 1076
QY 1077 PSEGAGSDVFDGDLGMAAGLQSLPHDPSPLQRYSEDTVPPLPSETDGYVAPLTCSPQ 1136
Db 1077 PSEGAGSDVFDGDLGMAAGLQSLPHDPSPLQRYSEDTVPPLPSETDGYVAPLTCSPQ 1136
QY 1137 PEYVNPQDVVRPQPPSPREGPLPAARPAAGATLERAKTILSPKNGVVKDVFAFGGAVENPEY 1196

Db 1137 PEYVQDVPRPQPSPREGPLPAARPAAGATLERPKTLSPGKNGVVDVFAFGAVENPEY 1196
QY 1197 LTPQGAAPQHPHPAFSPAFDNLVYWDQDPPRGAAPPSTFKGTPTAENPEYLGLDVPV 1255
Db 1197 LTPQGAAPQHPHPAFSPAFDNLVYWDQDPPRGAAPPSTFKGTPTAENPEYLGLDVPV 1255
RESULT 10
AAG88267
ID AAG88267 standard; Protein; 1255 AA.
XX AAG88267;
AC AAG88267;
XX
DT 11-SEP-2001 (first entry)
XX
DE HER2/neu amino acid sequence.
XX
KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
OS Homo sapiens.
XX
PN WO200141787-A1.
XX
PD 14-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US33591.
XX
PR 10-DEC-1999; 99US-0458299.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
PI Keogh E;
XX
DR WPI; 2001-374995/39.
XX
PT An isolated prepared HER2/neu epitope useful in a vaccine for inducing
cellular immune responses for the prevention and treatment of cancer -
XX
PS Disclosure; Page 15; 199pp; English.
XX
CC The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (I) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
CC and immunostimulant activities, and can be used in vaccines. (I), (II)
CC and (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample form a patient and
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour 'escape' due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention.
XX
SQ Sequence 1255 AA;

Query Match 98.3%; Score 6694; DB 22; Length 1255;

Best Local Similarity 98.3%; Pred. No. 0;
Matches 1237; Conservative 5; Mismatches 9; Indels 8; Gaps 2;
QY 1 MELAALCRWGLLALLPPGAASSTQVCTFNNFVSEFWLRVPKVSASHLE----LYQCQV 56
Db 1 MELAALCRWGLLALLPPGAASSTQVCTGDMK----LRLPASPEHDLMLRHLHYQCQV 56
QY 57 QGNLELTYPNTNASISFLQDIQEVQGYVLIHNVQVPLQRLRIVRGTQOLFEDNYALAV 116
Db 57 QGNLELTYPNTNASISFLQDIQEVQGYVLIHNVQVPLQRLRIVRGTQOLFEDNYALAV 116
QY 117 LDNGDPLNNTTPTVTGASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILWKDIFHN 176
Db 117 LDNGDPLNNTTPTVTGASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILWKDIFHN 176
QY 177 NOLALTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCC 236
Db 177 NOLALTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCC 236
QY 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
Db 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
QY 297 TACPYNILSTDVGSCTLVCPHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLREVRV 356
Db 297 TACPYNILSTDVGSCTLVCPHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLREVRV 356
QY 357 TSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFPETLEITGVLYISAMP 416
Db 357 TSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFPETLEITGVLYISAMP 416
QY 417 DSLPDLNVFQNLQVIRGRILHNGAYSLTLQGLIGISWLGSLRLSRELGLALIHNTLCLF 476
Db 417 DSLPDLNVFQNLQVIRGRILHNGAYSLTLQGLIGISWLGSLRLSRELGLALIHNTLCLF 476
QY 477 VHTVWDQLFRPHQALLHTANRDECEVGEGLACHOLCARGHCWGPGTQCNCQSFRLR 536
Db 477 VHTVWDQLFRPHQALLHTANRDECEVGEGLACHOLCARGHCWGPGTQCNCQSFRLR 536
QY 537 GQECVEECKVLQGLPREVYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPC 596
Db 537 GQECVEECKVLQGLPREVYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPC 596
QY 597 VARCPGKVPKELSYMPIWKFPEDEGACQPCINCTHSCVDLDDKGCAPQASPLTSIVS 656
Db 597 VARCPGKVPKELSYMPIWKFPEDEGACQPCINCTHSCVDLDDKGCAPQASPLTSIVS 656
QY 657 AVVGILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAOQNRILK 716
Db 657 AVVGILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAOQNRILK 716
QY 717 ETELKRVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEIILDEAYVMAG 776
Db 717 ETELKRVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEIILDEAYVMAG 776
QY 777 VGSVPSVRLGLCLTSTVOLVTLQMPYGLLDHVENRGRGLSQDILLNMCQIAKMSYL 836
Db 777 VGSVPSVRLGLCLTSTVOLVTLQMPYGLLDHVENRGRGLSQDILLNMCQIAKMSYL 836
QY 837 EDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEXHADGKGKVPKWMALLESILR 896
Db 837 EDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEXHADGKGKVPKWMALLESILR 896
QY 897 RFTHQSDVMSYGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVYIMV 956
Db 897 RFTHQSDVMSYGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVYIMV 956
QY 957 KCWMIIDSECRPFRELVSFSESMARDPQRFVVIQNEDELGPASPLDSTFYRSLEDDMDGD 1016
Db 957 KCWMIIDSECRPFRELVSFSESMARDPQRFVVIQNEDELGPASPLDSTFYRSLEDDMDGD 1016
QY 1017 LVDABEYLVPQOGFFCPDPAAGQGMVHRHRSSTRSGGDLTLGLFSEEPSPPLA 1076
Db 1017 LVDABEYLVPQOGFFCPDPAAGQGMVHRHRSSTRSGGDLTLGLFSEEPSPPLA 1076

Db 1017 LVDAEYLVQQGFCDDPAPAGAGMVHRRSSSTRSGGDLTLGLEPSEEAAPSPLA 1076
QY 1077 PSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1136
Db 1077 PSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1136
QY 1137 PEYVQPDVVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAFGGAVENPEY 1196
Db 1137 PEYVQPDVVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAFGGAVENPEY 1196
QY 1197 LTPQGAAPQHPHPPAPAFDNLYYWDQPPPERGAPSTFKGTPTAENPEYILGLDVPV 1255
Db 1197 LTPQGAAPQHPHPPAPAFDNLYYWDQPPPERGAPSTFKGTPTAENPEYILGLDVPV 1255

RESULT 11
AAE24067
ID AAE24067 standard; Protein; 1255 AA.
AC AAE24067;
XX
DT 23-SEP-2002 (first entry)
XX
DE Human Her-2 protein.
XX
KW Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KW hyperproliferative disorder; prophylaxis; inflammation; antisense;
KW tumour; gene therapy; phosphorothioate backbone.
XX
OS Homo sapiens.
XX
PN WO200222636-A1.
XX
PD 21-MAR-2002.
XX
PF 12-SEP-2001; 2001WO-US28572.
XX
PR 15-SEP-2000; 2000US-0663834.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Bennett CF, Cowser LM;
XX
DR WPI; 2002-471192/50.
DR N-PSDB; AAD38904.
XX
PT Novel antisense oligonucleotide which modulates the expression of Human
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT inflammation or to prevent infection in humans -
XX
PS Example 13; Page 95-107; 116pp; English.
XX
CC The invention relates to antisense compounds targetted to a nucleic
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC that specifically hybridises with and inhibits the expression of Her2.
CC Antisense compounds of the invention are used for treating diseases or
CC conditions associated with Her2 such as hyperproliferative disorders
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC neural or cardiac cancer. They are also useful prophylactically e.g.
CC to prevent or delay infection, inflammation and tumour formation. The
CC invention is also used in gene therapy. The present sequence is human
CC Her-2 protein.
XX
SQ Sequence 1255 AA;

Query Match 98.3%; Score 6694; DB 23; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1237; Conservative 5; Mismatches 9; Indels 8; Gaps 2;

QY 1 MELAALCRWGLLLALLPPGAASQVCTFNFTVSWLRVPKVSASHLE----LYQGCQV 56
Db 1 MELAALCRWGLLLALLPPGAASQVCTGDMK----LRUPASPETHLMLRHLYQGCQV 56

QY 57 QGNLELYLPTNASLSFLQDIOEVQGYVLIHAHQVRQVPLQRLIRIVRGTLQFEDNYALAV 116
Db 57 QGNLELYLPTNASLSFLQDIOEVQGYVLIHAHQVRQVPLQRLIRIVRGTLQFEDNYALAV 116
QY 117 LONGDPLNNTTPTVGTASPGGLRELQLRSUTEILKGGVLIQRNPOLCYQDTILWKDIFHKN 176
Db 117 LONGDPLNNTTPTVGTASPGGLRELQLRSUTEILKGGVLIQRNPOLCYQDTILWKDIFHKN 176
QY 177 NOLALTLIDTNSRACHPCSPMKSGRCWGESSEDCQSLTRTVTCAGGCARCKGPLPTDCC 236
Db 177 NOLALTLIDTNSRACHPCSPMKSGRCWGESSEDCQSLTRTVTCAGGCARCKGPLPTDCC 236
QY 237 HEQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
Db 237 HEQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
QY 297 TACPNYLTSDVGSCTLVCPHNEQVTAEDGTORCEKSKPCARVCYGLGMEHLRVRVAV 356
Db 297 TACPNYLTSDVGSCTLVCPHNEQVTAEDGTORCEKSKPCARVCYGLGMEHLRVRVAV 356
QY 357 TSANTIOEFAGCKKIFGSLAFLPESPDGDPASNTAPLQPELOVFFETLEITGYLYISAMP 416
Db 357 TSANTIOEFAGCKKIFGSLAFLPESPDGDPASNTAPLQPELOVFFETLEITGYLYISAMP 416
QY 417 DSIPLDSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLSGSLALIHNTLHLCF 476
Db 417 DSIPLDSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLSGSLALIHNTLHLCF 476
QY 477 VHTVPWDQLFRNPQOALLHTANRPDECEVGEGLACHQLCARGHCWGPGTQCVCNCSQFLR 536
Db 477 VHTVPWDQLFRNPQOALLHTANRPDECEVGEGLACHQLCARGHCWGPGTQCVCNCSQFLR 536
QY 537 GQECVVEECRVLOGLPREYVNAHCLPCHPECOPOGNSVTCFGEADQCACAHYKDPFPC 596
Db 537 GQECVVEECRVLOGLPREYVNAHCLPCHPECOPOGNSVTCFGEADQCACAHYKDPFPC 596
QY 597 VARCPGKVPDLSYMPKFWPDEEGACQPCPINCTHSCVDLDDKGCAPQASPLTSIVS 656
Db 597 VARCPGKVPDLSYMPKFWPDEEGACQPCPINCTHSCVDLDDKGCAPQASPLTSIVS 656
QY 657 AVVGILLVVVLGVVFGIILKRRQOKIRKYTMRLRLOETELVEPLTPSGAMPNQAQMRILK 716
Db 657 AVVGILLVVVLGVVFGIILKRRQOKIRKYTMRLRLOETELVEPLTPSGAMPNQAQMRILK 716
QY 717 ETELKRVKVLGSGAFCTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAVVMAG 776
Db 717 ETELKRVKVLGSGAFCTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAVVMAG 776
QY 777 VGSPYVSRLLGICLTSTVQLVTLQMLPYGCLLDHVRENRGRGLSGQDILLNMCQIAKMSYL 836
Db 777 VGSPYVSRLLGICLTSTVQLVTLQMLPYGCLLDHVRENRGRGLSGQDILLNMCQIAKMSYL 836
QY 837 EDVRLVHRDLAARNVLVKSNNHVKITDFGLABLLDIDETEHADGKGVPIKMALESILR 896
Db 837 EDVRLVHRDLAARNVLVKSNNHVKITDFGLABLLDIDETEHADGKGVPIKMALESILR 896
QY 897 RSPFHQSQVWSGVVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPPDICTIDVYIMV 956
Db 897 RSPFHQSQVWSGVVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPPDICTIDVYIMV 956
QY 957 KCWMDIDSECRPRFRELVSFSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMGD 1016
Db 957 KCWMDIDSECRPRFRELVSFSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMGD 1016
QY 1017 LVDABEYLVPOQGFCCPDPAAGAGMVHRRSSSTRSGGDLTLGLEPSEEAAPSPLA 1076
Db 1017 LVDABEYLVPOQGFCCPDPAAGAGMVHRRSSSTRSGGDLTLGLEPSEEAAPSPLA 1076
QY 1077 PSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1136
Db 1077 PSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQ 1136
QY 1137 PEYVQPDVVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAFGGAVENPEY 1196

Db 1137 PEYVNPQDVRPQPPSPREGPLPAARAGATLERPKTLSPGKNGVVKDVFAFGAVENPEY 1196
QY 1197 LTPQGGAAQPPHPPAFSPAFDNLYYWDQDPPPERGAPPSTFFKGTPTAENPEYLGLDVPV 1255
Db 1197 LTPQGGAAQPPHPPAFSPAFDNLYYWDQDPPPERGAPPSTFFKGTPTAENPEYLGLDVPV 1255
RESULT 12
AAE20479
ID AAE20479 standard; Protein; 1255 AA.
XX
AC AAE20479;
XX
DT 01-JUL-2002 (first entry)
XX
DE Human Her-2/neu protein.
XX
KW Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
KW human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 1021..1030
FT /note= "Naturally processed HLA-B44-restricted epitope"
XX
PN WO200214503-A2.
XX
PD 21-FEB-2002.
XX
PF 14-AUG-2001; 2001WO-US41733.
XX
PR 14-AUG-2000; 2000US-225152P.
PR 28-SEP-2000; 2000US-236428P.
PR 21-FEB-2001; 2001US-270520P.
XX
XX (CORI-) CORIXA CORP.
XX
PI Hand-zimmermann S, Cheever MA, Foy TW, Lodes MJ, Kalos MD;
PI McNeill PD, Vedvick TS;
XX
DR WPI; 2002-280758/32.
DR N-PSDB; AAD32743.
XX
PT Novel isolated Her-2/Neu polypeptide composition useful for therapy,
PT prevention and diagnosis of cancer, preferably breast cancer -
XX
PS Disclosure; Page 114-117; 129pp; English.
XX
CC The invention relates to an isolated Her-2/Neu polypeptide composition
CC effective for eliciting an immune response. The invention is useful for
CC eliciting an immune response in a patient, where the patient is human
CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
CC The composition is useful for the therapy and diagnosis of cancer,
CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
CC and other compositions for the diagnosis, prevention and treatment of
CC human malignancies, for stimulating and/or expanding T cells specific for
CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a
CC patient. The invention is useful for stimulating a T cell response in a
CC human patient, as probe or primer for nucleic acid hybridisation, to
CC selectively form duplex molecules with complementary stretches of the
CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full
CC length gene from a suitable library, and to direct expression of a
CC polypeptide in appropriate host cells. The composition is useful in
CC prophylactic or therapeutic applications and for the treatment of cancer,
CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-
CC associated malignancies. The invention is useful in gene therapy. The
CC present sequence is human Her-2/neu protein.
XX
SQ Sequence 1255 AA;

Query Match 98.3%; Score 6694; DB 23; Length 1255;

Best Local Similarity 98.3%; Pred. No. 0;
Matches 1237; Conservative 5; Mismatches 9; Indels 8; Gaps 2;
QY 1 MELAALCRWGLLALLPPGAASTQVCTFNNFVSVFWLRVPKVSASHLE----LYQCQCVV 56
Db 1 MELAALCRWGLLALLPPGAASTQVCTGDMK----LRLPASPETHDMLRHLHYQCQCVV 56
QY 57 QGNLELTYPNVAISFLQDIOEVQGYVLIJAHNQVRQVPLQRLRIVRGTLQFEDNVALAV 116
Db 57 QGNLELTYPNVAISFLQDIOEVQGYVLIJAHNQVRQVPLQRLRIVRGTLQFEDNVALAV 116
QY 117 LDNGDPLANNTPVTGASPGGLRELOLRSITELTKGGVLIQRPOLCYQDTILWKDIFHKN 176
Db 117 LDNGDPLANNTPVTGASPGGLRELOLRSITELTKGGVLIQRPOLCYQDTILWKDIFHKN 176
QY 177 NQLALTLIDTNRSRACHPCSPMKGSRGWGESSEDCQSLTRTVCAAGGACRCKGPLTDC 236
Db 177 NQLALTLIDTNRSRACHPCSPMKGSRGWGESSEDCQSLTRTVCAAGGACRCKGPLTDC 236
QY 237 HEQCAAGCTGPIGHSCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
Db 237 HEQCAAGCTGPIGHSCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
QY 297 TACPYNYLSTDVGSCTLYCPLHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLREVAV 356
Db 297 TACPYNYLSTDVGSCTLYCPLHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLREVAV 356
QY 357 TSANIQEFAGCCKIFGSLAFLPESFDGPPASNTAPLQPEQLQVFTLEITGYLYISAMP 416
Db 357 TSANIQEFAGCCKIFGSLAFLPESFDGPPASNTAPLQPEQLQVFTLEITGYLYISAMP 416
QY 417 DSLPDLVSFQNLQVIRGRILHNGAYSITLQGLIGISWGLRSURELSGLALIHNNTHLCF 476
Db 417 DSLPDLVSFQNLQVIRGRILHNGAYSITLQGLIGISWGLRSURELSGLALIHNNTHLCF 476
QY 477 VHTVPWDQLFRPHQALLHTANRPEDECVEGLACHOLCARGHCVGPPTQCVNCSQFLR 536
Db 477 VHTVPWDQLFRPHQALLHTANRPEDECVEGLACHOLCARGHCVGPPTQCVNCSQFLR 536
QY 537 GOECVEECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFPGPADQCVACAHYKDPFPC 596
Db 537 GOECVEECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFPGPADQCVACAHYKDPFPC 596
QY 597 VARCPGKVPDLVSYPIMKFPDEEGACQPCPINCTHSCVDLDDKCPABORASPLTSIVS 656
Db 597 VARCPGKVPDLVSYPIMKFPDEEGACQPCPINCTHSCVDLDDKCPABORASPLTSIVS 656
QY 657 AVVGILLVVVLGVWFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILK 716
Db 657 AVVGILLVVVLGVWFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILK 716
QY 717 ETELKRVKVLGSGAFGTGVVKGWI PDGENVKIPVAIKVIRENTSPKANKEIIDEAVVMAG 776
Db 717 ETELKRVKVLGSGAFGTGVVKGWI PDGENVKIPVAIKVIRENTSPKANKEIIDEAVVMAG 776
QY 777 VGSPVSVRLIGLICLTSTVQLVTQLMPYGLLDHVRENRRGLSGQDLNWCMIAGMSYL 836
Db 777 VGSPVSVRLIGLICLTSTVQLVTQLMPYGLLDHVRENRRGLSGQDLNWCMIAGMSYL 836
QY 837 EDVRLVHRDLAARNVLKSPNHNKLTDFGLARLLDIDETEHADGGKVPKIMWALESILR 896
Db 837 EDVRLVHRDLAARNVLKSPNHNKLTDFGLARLLDIDETEHADGGKVPKIMWALESILR 896
QY 897 RRFTHQSDVMSVGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITDVMYIMV 956
Db 897 RRFTHQSDVMSVGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITDVMYIMV 956
QY 957 KCWMIIDSECRPRFRELVSFSESMARDPQRFVVIQNEDELGPASPLDSTFYRSLLEDDMGD 1016
Db 957 KCWMIIDSECRPRFRELVSFSESMARDPQRFVVIQNEDELGPASPLDSTFYRSLLEDDMGD 1016
QY 1017 LVDASEYILVPQOGFFCPDPAAGAGMWHHRSSSTRSGGDLTLGLSPSESEAPRPLA 1076
Db 1017 LVDASEYILVPQOGFFCPDPAAGAGMWHHRSSSTRSGGDLTLGLSPSESEAPRPLA 1076

Db 1017 LVDAEYLVQPFCCPDPAAGAGVHHRSSSTRSGGDLTLGLEPSEEARSPPLA 1076
QY 1077 PSEGAGSDVFDGLGMAAGLQSLPHDPSPLQRYSEDTVPPLPSETDGYVAPLTCSPQ 1136
Db 1077 PSEGAGSDVFDGLGMAAGLQSLPHDPSPLQRYSEDTVPPLPSETDGYVAPLTCSPQ 1136
QY 1137 PEYVNPQDVPRQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVFAFGGAVENPEY 1196
Db 1137 PEYVNPQDVPRQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVFAFGGAVENPEY 1196
QY 1197 LTPQGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
Db 1197 LTPQGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 13

AA051143
ID AA051143 standard; Protein; 1255 AA.

AC AA051143;

XX 17-JUN-2002 (first entry)

XX Human Her-2/neu oncogene-encoded p185 glycoprotein.

DE Her-2/neu; oncogene; cancer; tumor; vaccine; human; p185;
KW tyrosine kinase; receptor; c-erbB2; gene therapy.

XX Homo sapiens.

XX Key Location/Qualifiers
FH Domain 1..653
FT Domain /note= "extracellular domain"
FT Domain 676..1255
FT Domain /note= "intracellular domain"
FT Domain 990..1255
FT Domain /note= "phosphorylation domain"

XX WO200212341-A2.

XX 14-FEB-2002.

XX 03-AUG-2001; 2001WO-US24283.

XX 03-AUG-2000; 2000US-0632507.

XX (CORI-) CORIXA CORP.
XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Cheever MA, Gheysen D;

XX WPI; 2002-241743/29.

XX N-PSDB; ABA92250.

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting
PT or enhancing an immune response to the protein, has Her-2/neu
PT extracellular domain fused to Her-2/neu intracellular or
PT phosphorylation domain

XX Claim 68; Fig 7; 141pp; English.

XX The present sequence is that of human Her-2/neu (p185 glycoprotein
CC or c-erbB2), an oncogenic self-protein and target for anti-cancer
CC vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
CC in a variety of cancers, including breast, ovarian, colon, lung and
CC prostate cancer. Her-2/neu is a member of the tyrosine kinase
CC family of receptor-like glycoproteins. It comprises an extracellular
CC domain with homology to the epidermal growth factor receptor
CC (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
CC intracellular domain that also shows homology to EGFR. Its
CC overexpression correlates with a poor prognosis in breast and
CC ovarian cancers. The invention provides Her-2/neu fusion
CC proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In
CC preferred fusion proteins, the extracellular domain of a Her-2/neu
CC protein is fused to a Her-2/neu intracellular domain or
CC phosphorylation domain (or its DeltapB fragment). An immune
CC response to Her-2/neu protein is elicited or enhanced by
CC administering the fusion protein in the form of a vaccine, or by
CC transfecting cells of an animal ex vivo with a nucleic acid
CC encoding the fusion protein, and delivering the transfected cells
CC to the animal. The fusion proteins, nucleic acids, and isolated
CC specific T-cells are useful for inhibiting the development of a
CC cancer, especially breast, ovarian, colon, lung or prostate cancer
CC in a patient. T cells that specifically react with a Her-2/neu
CC fusion protein can be used to remove tumour cells from a sample in
CC order to inhibit the development of cancer in a patient.

XX Sequence 1255 AA;

Query Match 98.3%; Score 6694; DB 23; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1237; Conservative 5; Mismatches 9; Indels 8; Gaps 2;

QY 1 MELAALCRWGLLLALLPPGAASTQVCTFNNFTVSFWLRVPKVSASHLE----LYQCQV 56
Db 1 MELAALCRWGLLLALLPPGAASTQVCTGTDK----LRLPASPETHLDMRLHLYQCQV 56
QY 57 QGNLELTYPNTWASLSFLQDIOEVQYVLIANOVQVPLQRLIRVRGTQLEDNVALAV 116
Db 57 QGNLELTYPNTWASLSFLQDIOEVQYVLIANOVQVPLQRLIRVRGTQLEDNVALAV 116
QY 117 LDNGDPLNNTTPTVTGASPGGLRELQRLSLTEILKGVLIQRNPOLCYQDTILWKDIFHKN 176
Db 117 LDNGDPLNNTTPTVTGASPGGLRELQRLSLTEILKGVLIQRNPOLCYQDTILWKDIFHKN 176
QY 177 NQALTLTDNRSRACHPCSPCKSGCWGESSEDCQSLTRTVCCAGCARCGPLTDCC 236
Db 177 NQALTLTDNRSRACHPCSPCKSGCWGESSEDCQSLTRTVCCAGCARCGPLTDCC 236
QY 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESNPNEGRTYTCASCV 296
Db 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESNPNEGRTYTCASCV 296
QY 297 TACPYNLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVAV 356
Db 297 TACPYNLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVAV 356
QY 357 TSANTQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEITGYLISAWP 416
Db 357 TSANTQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEITGYLISAWP 416
QY 417 DSLPDLSPFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLGSLALHNNTHLCF 476
Db 417 DSLPDLSPFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLGSLALHNNTHLCF 476
QY 477 VHTVPMQDLFRNPQALLHTANRPEDECVGEGLACHQLCARGCHGPGPTQCVCNSQFLR 536
Db 477 VHTVPMQDLFRNPQALLHTANRPEDECVGEGLACHQLCARGCHGPGPTQCVCNSQFLR 536
QY 537 GOECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPC 596
Db 537 GOECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPC 596
QY 597 VARCPGVKPDLSYMPINWKFDEEGACOPCPINCTHSCVDLDDKGPAPORASPLTSIVS 656
Db 597 VARCPGVKPDLSYMPINWKFDEEGACOPCPINCTHSCVDLDDKGPAPORASPLTSIVS 656
QY 657 AVVGILLVVVLGVWFGLIKRQQKIRKYTWRLLEQETELVEPLTPSGAMPNQAMRILK 716
Db 657 AVVGILLVVVLGVWFGLIKRQQKIRKYTWRLLEQETELVEPLTPSGAMPNQAMRILK 716
QY 717 ETELKRVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAG 776
Db 717 ETELKRVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAG 776

QY 777 VGSPPYVSRLLGICLTSTVQLVLTQMPYGCCLLDHVHRENRGRGLSGQDILLNWCQIAKMSYL 836
Db 777 VGSPPYVSRLLGICLTSTVQLVLTQMPYGCCLLDHVHRENRGRGLSGQDILLNWCQIAKMSYL 836
QY 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILR 896
Db 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILR 896
QY 897 RRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPQPICTIDVYMIW 956
Db 897 RRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPQPICTIDVYMIW 956
QY 957 KCMWIDSECPRELFSEFSRMARDQRFVWQNEIDLGPASPLDSTFYRSILLEDDMGD 1016
Db 957 KCMWIDSECPRELFSEFSRMARDQRFVWQNEIDLGPASPLDSTFYRSILLEDDMGD 1016
QY 1017 LVDAEYLVPOQGFCDPAPGAGGVMHRRHSSTRSGGDLTTLGLEPSEEAAPSPPLA 1076
Db 1017 LVDAEYLVPOQGFCDPAPGAGGVMHRRHSSTRSGGDLTTLGLEPSEEAAPSPPLA 1076
QY 1077 PSEGAGSDVFDGLGMAAGLQSLPTHDPSPQLQYSEDPTVPLPSETDGYVAPLTCSPQ 1136
Db 1077 PSEGAGSDVFDGLGMAAGLQSLPTHDPSPQLQYSEDPTVPLPSETDGYVAPLTCSPQ 1136
QY 1137 PEYVNOQDVRPQPPSPREGPLPAARFAGATLERAKTILSPKNGVWVDVFAFGAVENPEY 1196
Db 1137 PEYVNOQDVRPQPPSPREGPLPAARFAGATLERAKTILSPKNGVWVDVFAFGAVENPEY 1196
QY 1197 LTPQGGAAPOPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGPTAENPEYLGLOVPV 1255
Db 1197 LTPQGGAAPOPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGPTAENPEYLGLOVPV 1255

RESULT 14

AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX
AC AAU77114;
XX
DT 05-JUN-2002 (first entry)
XX
DE Human Her-2/neu polypeptide.
XX
KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AMI; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
XX
OS Homo sapiens.
XX
PN WO200213847-A2.
XX
PD 21-FEB-2002.
XX
PF 13-AUG-2001; 2001WO-US25408.
XX
PR 14-AUG-2000; 2000US-0638280.
PR 28-SEP-2000; 2000US-0675904.
XX
XX (CORI-) CORIXA CORP.
XX
XX Gaiger A, Cheever MA, Hand-zimmermann S;
PI WPI; 2002-280741/32.
DR N-PSDB; ABK10730.
DR
XX Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide
XX
XX Disclosure; Page 71-74; 74pp; English.
XX

CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;

Query Match 98.3%; Score 6694; DB 23; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1237; Conservative 5; Mismatches 9; Indels 8; Gaps 2;
QY 1 MELAALCRWGLLLALLPPGAASSTVCTFNNFTVSFWMRVKVSASHLE---LYQSCQVV 56
Db 1 MELAALCRWGLLLALLPPGAASSTVCTGTDMK----LRLPASPETHLDMRLHLYQSCQVV 56
QY 57 QGNLELYLPTNASLSFLQDIOEVQGYVLIHAHQVRQVPLQRLIRVRGTQLFEDNYALAV 116
Db 57 QGNLELYLPTNASLSFLQDIOEVQGYVLIHAHQVRQVPLQRLIRVRGTQLFEDNYALAV 116
QY 117 LONGDPLNNTTPTVGTASPGGLRELQRLSRLTEILKGVLIQRPOLCYQDTILWKDIFHKN 176
Db 117 LONGDPLNNTTPTVGTASPGGLRELQRLSRLTEILKGVLIQRPOLCYQDTILWKDIFHKN 176
QY 177 NOLALTLIDTNSRACHPCSPKSGRCWGESSEDCQSLTRTVCAGGCARCKGLPTDCC 236
Db 177 NOLALTLIDTNSRACHPCSPKSGRCWGESSEDCQSLTRTVCAGGCARCKGLPTDCC 236
QY 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNEPGRYTFGASCV 296
Db 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNEPGRYTFGASCV 296
QY 297 TACPNYLTSTVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVAV 356
Db 297 TACPNYLTSTVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVAV 356
QY 357 TSANTQEFAGCKKIFGSLAFLESFDGDPASNTAPLOPELOVFFETLEITGYLISAMP 416
Db 357 TSANTQEFAGCKKIFGSLAFLESFDGDPASNTAPLOPELOVFFETLEITGYLISAMP 416
QY 417 DSLPDLSPVQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLSGSLALHNNTHLCF 476
Db 417 DSLPDLSPVQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLSGSLALHNNTHLCF 476
QY 477 VHTVPWDQLFRNPHQALLHTANRPEDECYEGELACHQLCARGHCWGPPTQCVNCSQFLR 536
Db 477 VHTVPWDQLFRNPHQALLHTANRPEDECYEGELACHQLCARGHCWGPPTQCVNCSQFLR 536
QY 537 GOECVEECRVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFPC 596
Db 537 GOECVEECRVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFPC 596
QY 597 VARCPGKVPULSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIVS 656
Db 597 VARCPGKVPULSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIVS 656
QY 657 AVVGILLVVVLGVVFGILIKRQOKIRKYTMRLRLQETELVEPLTPSGAMPNQAMRILK 716
Db 657 AVVGILLVVVLGVVFGILIKRQOKIRKYTMRLRLQETELVEPLTPSGAMPNQAMRILK 716
QY 717 ETELKRVKVLGSGAGTGVYKGIWIDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAG 776
Db 717 ETELKRVKVLGSGAGTGVYKGIWIDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAG 776
QY 777 VGSPPVSRLLGICLTSTVQLVLTQMPYGCCLLDHVHRENRGRGLSGQDILLNWCQIAKMSYL 836
Db 777 VGSPPVSRLLGICLTSTVQLVLTQMPYGCCLLDHVHRENRGRGLSGQDILLNWCQIAKMSYL 836
QY 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILR 896

Db 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILR 896
QY 897 RRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYMW 956
Db 897 RRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYMW 956
QY 957 KCMWIDSECRPRELVSSEFSESRMARDPQRFVVIQNEEDLGSPASPLDSTFYRSLEDDDDMGD 1016
Db 957 KCMWIDSECRPRELVSSEFSESRMARDPQRFVVIQNEEDLGSPASPLDSTFYRSLEDDDDMGD 1016
QY 1017 LVDAEYLVPOQGFPCDPAPAGAGWVHHHRSSSTRSGGDLTLGLEPSEEBAPRSPLA 1076
Db 1017 LVDAEYLVPOQGFPCDPAPAGAGWVHHHRSSSTRSGGDLTLGLEPSEEBAPRSPLA 1076
QY 1077 PSEGAGSDVFDGLGMAAGKLOSLPHDPSPLQRYSEDTVPPLPSETDGYVAPLTCSPQ 1136
Db 1077 PSEGAGSDVFDGLGMAAGKLOSLPHDPSPLQRYSEDTVPPLPSETDGYVAPLTCSPQ 1136
QY 1137 PEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGWVKDVFAGGAVENPEY 1196
Db 1137 PEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGWVKDVFAGGAVENPEY 1196
QY 1197 LTPQGGAAPOHPHPPAFSPADNLYYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
Db 1197 LTPQGGAAPOHPHPPAFSPADNLYYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 15

AAR39568
ID AAR39568 standard; Protein; 1433 AA.

XX AAR39568;

AC AAR39568;

DT 07-FEB-1994 (first entry)

DE Sequence of c-erbB-2 tumour antigen.

XX Tumour antigen; c-erbB-2; glycoprotein.

XX Homo sapiens.

XX WO9316185-A.

XX 19-AUG-1993.

XX 05-FEB-1993; 93WO-US01055.

XX 06-FEB-1992; 92US-0831967.

XX (CETU) CETUS ONCOLOGY CORP.

XX (CREA-) CREATIVE BIOMOLECULES INC.

XX Houston LL, Huston JS, Oppermann H, Ring DB;

XX WPI; 1993-272889/34.

XX N-PSDB; AAQ46083.

XX New single chain Fv polypeptide binding to C-erbB-2 tumour

XX antigen - for imaging or treating breast or ovarian cancer etc.

XX Disclosure; pages 48-54; 87pp; English.

XX c-erbB-2 refers to a protein antigen expressed on the surface of

XX tumour cells, such as breast and ovarian tumour cells, which is an

XX approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric

XX pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39568 represents

XX the location of a stop codon in AAQ46083.

XX Sequence 1433 AA;

XX Query Match 97.8%; Score 6651; DB 14; Length 1433;

XX Best Local Similarity 97.7%; Pred. No. 0;

Matches 1230; Conservative 7; Mismatches 14; Indels 8; Gaps 2;
QY 1 MELAALCRWGLLLALLPPGAASQTCTFNNFTVSEFWLRVPKVSASHLE---LYQCQOVV 56
Db 1 MELAALCRWGLLLALLPPGAASQTCTGTDMK----LRLPASPEHDLMLRHLHYQVCOVV 56
QY 57 QGNLELYLPTNASLFLQDIOEVQGYVLIAHNQVRQVPLQRLRIVRGTLFEDNYALAV 116
Db 57 QGNLELYLPTNASLFLQDIOEVQGYVLIAHNQVRQVPLQRLRIVRGTLFEDNYALAV 116
QY 117 LQNGDPLNNTTVPVTCASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKN 176
Db 117 LQNGDPLNNTTVPVTCASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKN 176
QY 177 NOLAUTLIDTNSRACHPCSPKCSRCWGESSEDCQSLTRTVACAGGCARCKPLPTDCC 236
Db 177 NOLAUTLIDTNSRACHPCSPKCSRCWGESSEDCQSLTRTVACAGGCARCKPLPTDCC 236
QY 237 HEQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
Db 237 HEQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
QY 297 TACPNYILSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVAV 356
Db 297 TACPNYILSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVAV 356
QY 357 TSANIQEFAGCKKIFGSLAFPLPESPDGPASNTAPLOPELOQVFFETLEBITGYLYISAMP 416
Db 357 TSANIQEFAGCKKIFGSLAFPLPESPDGPASNTAPLOPELOQVFFETLEBITGYLYISAMP 416
QY 417 DSLPDLVSFQNLQVIRGRILHNGAYSLLTQGLGISWLGSLRSLRELGSGLALIHNNHLSF 476
Db 417 DSLPDLVSFQNLQVIRGRILHNGAYSLLTQGLGISWLGSLRSLRELGSGLALIHNNHLSF 476
QY 477 VHTVPWDLFRNPHOALLHTANRDECEVGEGLACHQLCARGHGWPGTQCNCQOFLR 536
Db 477 VHTVPWDLFRNPHOALLHTANRDECEVGEGLACHQLCARGHGWPGTQCNCQOFLR 536
QY 537 GOECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPC 596
Db 537 GOECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPC 596
QY 597 VARCPGSKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCAPQASPLTSIYS 656
Db 597 VARCPGSKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCAPQASPLTSIYS 656
QY 657 AVVGILLVVVLGVVFGILIKRRQKIRKYTMRRLLOETELVEPLTPSGAMPNQAQNRILK 716
Db 657 AVVGILLVVVLGVVFGILIKRRQKIRKYTMRRLLOETELVEPLTPSGAMPNQAQNRILK 716
QY 717 ETELKRVKVLGSGAFGTGVIWIPDGENVKIPVAIKVLRNTPSKANKEIIDEAYVMAG 776
Db 717 ETELKRVKVLGSGAFGTGVIWIPDGENVKIPVAIKVLRNTPSKANKEIIDEAYVMAG 776
QY 777 VGSPYVSRLLGICLTSTVQLVTQLMPYGLLDHVRNRLGSLQDLLNWCMIAGMSYL 836
Db 777 VGSPYVSRLLGICLTSTVQLVTQLMPYGLLDHVRNRLGSLQDLLNWCMIAGMSYL 836
QY 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILR 896
Db 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILR 896
QY 897 RRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYMW 956
Db 897 RRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYMW 956
QY 957 KCMWIDSECRPRELVSSEFSESRMARDPQRFVVIQNEEDLGSPASPLDSTFYRSLEDDDDMGD 1016
Db 957 KCMWIDSECRPRELVSSEFSESRMARDPQRFVVIQNEEDLGSPASPLDSTFYRSLEDDDDMGD 1016
QY 1017 LVDAEYLVPOQGFPCDPAPAGAGWVHHHRSSSTRSGGDLTLGLEPSEEBAPRSPLA 1076
Db 1017 LVDAEYLVPOQGFPCDPAPAGAGWVHHHRSSSTRSGGDLTLGLEPSEEBAPRSPLA 1076

Qy	1077	PSEGAGSDVFDGLGMAAKGLOSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTCSPQ	1136
Db	1077	PSEGAGSDVFDGLGMAAKGLOSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTCSPQ	1136
Qy	1137	PEYVNOQDVVRPQPSREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEY	1196
Db	1137	PEYVNOQDVVRPQPSREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEY	1196
Qy	1197	LTPQGGAAPQHPHPPAFSPAFDNLYYWDQDPPPERGAPPTFKGTPTAENPEYLGLDVPV	1255
Db	1197	LTPQGGAAPQHPHPPAFSPAFDNLYYWDQDPPPERGAPPTFKGTPTAENPEYLGLDVPV	1255

Search completed: July 22, 2003, 08:40:12
 Job time : 43.9774 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 23.0157 Seconds
(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-59-73-14
Perfect score: 6847
Sequence: 1 MEALACRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6707	98.0	1255	1 A24571	protein-tyrosine k
2	5901.5	86.2	1254	2 I48161	p-185 precursor -
3	5901	86.2	1260	1 TVRTNU	protein-tyrosine k
4	3098	45.2	1210	1 GQHUE	epidermal growth f
5	3074	44.9	1210	2 A53183	epidermal growth f
6	3054.5	44.6	1223	1 TVCHLV	epidermal growth f
7	2930.5	42.8	1308	2 A47253	epidermal growth f
8	2633	38.5	1166	1 S06142	protein-tyrosine k
9	2373.5	34.7	1342	2 A36223	kinase-related tra
10	2299.5	33.6	1339	2 JC4387	epidermal growth f
11	1786.5	25.8	698	1 TVFVLV	protein-tyrosine k
12	1703	24.9	604	1 TVYUHV	protein-tyrosine k
13	1650.5	24.1	1330	1 GQFFE	epidermal growth f
14	1647	24.1	544	2 S35745	protein-tyrosine k
15	1640	24.0	545	2 S00727	kinase-related tra
16	1623	23.7	540	2 B44776	protein-tyrosine k
17	1621	23.7	540	1 TVFVVB	protein-tyrosine k
18	1466	21.4	644	2 A36325	epidermal growth f
19	1274	18.6	1323	2 E88257	protein let-23 [im
20	1274	18.6	1374	2 S70712	protein-tyrosine k
21	1186	17.3	1369	2 S70713	protein-tyrosine k
22	1121	16.4	1717	1 A45558	epidermal growth f
23	1086	15.9	527	2 A42032	epidermal growth f
24	942.5	13.8	843	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	protein-tyrosine k
26	754.5	11.0	311	2 S13808	protein-tyrosine k
27	736	10.7	1363	2 T43220	insulin-like growth
28	720	10.5	1382	1 INHUR	insulin receptor p
29	713	10.4	1383	2 A36080	insulin receptor p

RESULT 1

A24571

Protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N;Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein erbB
C;Species: Homo sapiens (man)
C;Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999

C;Accession: A24571; A25491; A44188; B44188; I59509; I57622

R;Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;

Nature 319, 230-234, 1986

A;Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth f

A;Reference number: A24571; MUID:86118663; PMID:3003577

A;Accession: A24571

A;Molecule type: mRNA

A;Residues: 1-1255 <YAM>

A;Cross-references: GB:X03363; NID:G31197; PIDN:CAA27060.1; PID:G31198

R;Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.

Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A;Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epiderm

A;Reference number: A25491; MUID:86016729; PMID:2995967

A;Accession: A25491

A;Molecule type: DNA

A;Residues: 737-1031 <SEM>

A;Cross-references: GB:M11767; NID:G182163; PIDN:AAA35808.1; PID:G553282

R;Cousseus, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg, P.

Science 230, 1132-1139, 1985

A;Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromos

A;Reference number: A44188; MUID:86070181; PMID:2999974

A;Accession: A44188

A;Molecule type: DNA

A;Residues: 740-910 <COU1>

A;Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989

A;Accession: B44188

A;Molecule type: mRNA

A;Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>

A;Cross-references: GB:M11730; NID:G183986

R;King, C.R.; Kraus, M.H.; Aaronson, S.A.

Science 229, 974-976, 1985

A;Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A;Reference number: I59509; MUID:85272597; PMID:2992089

A;Accession: I59509

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 832-909 <REX>

A;Cross-references: GB:L29395; NID:G459807; PIDN:AAA35809.1; PID:G459808

R;Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987

A;Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptional

A;Reference number: I57622; MUID:87286898; PMID:3039351

A;Accession: I57622

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-191 <TAL>

A:Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C:Genetics:
A:Gene: GDB:ERBB2; NGL; NEU; HER-2
A:Cross-references: GDB:120613; OMIM:164870
A:Map position: 17q21.1-17q21.1
A:Introns: 25/1; 75/3; 147/1; 883/3
A:Note: the list of introns is incomplete
C:Function:
A:Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
inase
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F:22-653/Domain: extracellular #status predicted <EXT>
F:70-304/Domain: EGF receptor extracellular domain repeat <EB1>
F:395-605/Domain: EGF receptor extracellular domain repeat <EB2>
F:654-675/Domain: transmembrane #status predicted <TM>
F:676-1255/Domain: intracellular #status predicted <INT>
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif
F:68,124,187,259,530,571,629/Binding site: carboxydrate (Asn) (covalent) #status predicted
F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:753/Active site: Lys #status predicted
F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 98.0%; Score 6707; DB 1; Length 1255;
Best Local Similarity 98.3%; Pred. No. 2.9e-267;
Matches 1240; Conservative 5; Mismatches 10; Indels 6; Gaps 2;

Qy	1	MELAAICRWGLLLALLPPGAASCTGCTDMKRLPASPETHDMLRHLVGGCQVQGNL	60
Db	1	MELAAICRWGLLLALLPPGAASCTGCTDMKRLPASPETHDMLRHLVGGCQVQGNL	60
Qy	61	ELTYLPTNASLSFLQDIEQVFNFTVSFWLRVPKVSASHLEQRLRIVRGTLQFEDNYAL	120
Db	61	ELTYLPTNASLSFLQDIEQV- -GVLIHQVQVPL- - -QRLRIVRGTLQFEDNYAL	114
Qy	121	AVLNGDPLNNTPTVTGASPGGLRELQRLSTEILKGGVLIQRNPOLCYQDTILWKDIFH	180
Db	115	AVLNGDPLNNTPTVTGASPGGLRELQRLSTEILKGGVLIQRNPOLCYQDTILWKDIFH	174
Qy	181	KNQALATLIDNRSRACHPCSPCKGSRGWGSESSDCQSLTRTVACGGCARCKGPLPTD	240
Db	175	KNQALATLIDNRSRACHPCSPCKGSRGWGSESSDCQSLTRTVACGGCARCKGPLPTD	234
Qy	241	CCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGAS	300
Db	235	CCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGAS	294
Qy	301	CVTACPNYLTVDVGSCTLVPLHNQEVTAEDGTQRCBKSKPCARVCYGLGMHLEVR	360
Db	295	CVTACPNYLTVDVGSCTLVPLHNQEVTAEDGTQRCBKSKPCARVCYGLGMHLEVR	354
Qy	361	AVTSANTQEFAGCKKIFGSLAFLESFDGDPASNTAPLOPEQLQVFLEITGYLYISA	420
Db	355	AVTSANTQEFAGCKKIFGSLAFLESFDGDPASNTAPLOPEQLQVFLEITGYLYISA	414
Qy	421	WPDLSPLDSVFQNLQVIRGRILHNGAYSLLTLQGLGISWLGRLSRLRELGSGLALIHNTHL	480
Db	415	WPDLSPLDSVFQNLQVIRGRILHNGAYSLLTLQGLGISWLGRLSRLRELGSGLALIHNTHL	474
Qy	481	CFVHTVPWDQLFRPHQALLTANRDECECVGEGGLACHQLCARGHCWGPGTQCVCNSQF	540
Db	475	CFVHTVPWDQLFRPHQALLTANRDECECVGEGGLACHQLCARGHCWGPGTQCVCNSQF	534
Qy	541	LRGQECVEECVLOGLPREYVNAHCLPCHPECOPNGSVTCFCPEADQCACAHYKDDP	600
Db	535	LRGQECVEECVLOGLPREYVNAHCLPCHPECOPNGSVTCFCPEADQCACAHYKDDP	594
Qy	601	FCVARCPGKPDLSYMPINWKFPEDEGACQPCPINCTHSCVDLDDKGPAPQASPLTSI	660

Db	595	FCVARCPGKPDLSYMPINWKFPEDEGACQPCPINCTHSCVDLDDKGPAPQASPLTSI	654
Qy	661	VSAVWGILLVVLGVVFGIILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAWRI	720
Db	655	ISAVWGILLVVLGVVFGIILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAWRI	714
Qy	721	LKETELRKVVLGSGAGFTGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVM	780
Db	715	LKETELRKVVLGSGAGFTGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVM	774
Qy	781	AGVSPVSRLLGICLTSTVOLVQLMPYCCLLDHVRNRLGRLGSDLLNWCQIAKMS	840
Db	775	AGVSPVSRLLGICLTSTVOLVQLMPYCCLLDHVRNRLGRLGSDLLNWCQIAKMS	834
Qy	841	YLEDVRLVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGKGKPIKMALESI	900
Db	835	YLEDVRLVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGKGKPIKMALESI	894
Qy	901	LRRRFTHQSDVWSVGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVMI	960
Db	895	LRRRFTHQSDVWSVGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVMI	954
Qy	961	MVXCMWIDSECRPRFRELVSFSEFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDM	1020
Db	955	MVXCMWIDSECRPRFRELVSFSEFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDM	1014
Qy	1021	GDLVDABEYLVPQGGFCPPDPAPGAGGVVHRRSSSTRSGGDLTLGLPSEEEARSP	1080
Db	1015	GDLVDABEYLVPQGGFCPPDPAPGAGGVVHRRSSSTRSGGDLTLGLPSEEEARSP	1074
Qy	1081	LAPSEAGSDVDFDGLGMAAKGLQSLPHTDPSPLOQYSEDPTVPLPSETDGYVAPLTCS	1140
Db	1075	LAPSEAGSDVDFDGLGMAAKGLQSLPHTDPSPLOQYSEDPTVPLPSETDGYVAPLTCS	1134
Qy	1141	POPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLPFGKXGNGVVKDVFAGGAVENP	1200
Db	1135	POPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLPFGKXGNGVVKDVFAGGAVENP	1194
Qy	1201	EYLTPOGGAAPQHPHPPAPSPADNLYYWDODPPERCAPSTFKGTPTAENPEVLGDVP	1260
Db	1195	EYLTPOGGAAPQHPHPPAPSPADNLYYWDODPPERCAPSTFKGTPTAENPEVLGDVP	1254
Qy	1261	V 1261	
Db	1255	V 1255	

RESULT 2
I48161
P-185 precursor - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C:Accession: I48161
R:Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa, Gene 140, 251-255, 1994
A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A:Reference number: I48161; MUID:94193007; PMID:7908275
A:Accession: I48161
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1254 <RES>
A:Cross-references: GB:D16295; NID:g493236; PIDN:BAA03801.1; PID:g747595
C:Genetics:
A:Gene: neu
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif

Query Match 86.2%; Score 5901.5; DB 2; Length 1254;
Best Local Similarity 86.2%; Pred. No. 2.5e-234;
Matches 1088; Conservative 60; Mismatches 105; Indels 9; Gaps 3;

Qy 1 MELAALCRWGLLLALLPPGAASSTVCTGTDMMKRLPASPTHLDMLRHLHYGCGQVVOGNL 60
Db 1 MELAAWCGWGLLLALLSPGASGTCTGTDMMKRLPASPTHLDIVRHLHYGCGQVVOGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEQVFNFTVSFMLRVPKVSASHLE-ORLRIVRGTLPEONYA 119
Db 61 ELTYLPANATLSFLQDIQEQV-----GYMLIAHSQVRHVPLQRLRIVRGTLPEDKYA 113
Qy 120 LAVLDNGDPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIF 179
Db 114 LAVLDNRDPLNVTATGRTPEGLRELQRLSLTEILKGGVLIQRNPOLCYQDTVLWKDVP 173
Qy 180 HKNNQLALTLIDNRSRACHPCSPMKGSRCWGESSEDCQSILRTVACGGCARCKGLPT 239
Db 174 RKNQLAPVDITNRSRACPPCAPACKDNHCWGPEDCQTLTGTTIAPRAVPAARARLPT 233
Qy 240 DCHEQCAAGCTGPKHSDCLACLPNHSIGICELHCPALVTYNTDTFESMPNPEGRYTFGA 299
Db 234 DCHEQCAAGCTGPKHSDCLACLPNHSIGICELHCPALVTYNTDTFESMPNPEGRYTFGA 293
Qy 300 SCVTACPNYLSLTDVGSCTVLCPLHNOEVTAAEDGTQRCCKSKPCARVCYGLGMEHLREV 359
Db 294 SCVTTCPNYLSLTDVGSCTVLCPLNNOEVTAAEDGTQRCCKSKPCARVCYGLGMEHLRGA 353
Qy 360 RAVTSANIQEPAGCKKIFGSLAFIPESPFGDPDPAANTAPLOEQLOVFTETLEETGYLYIS 419
Db 354 RAVTSANIQEPAGCKKIFGSLAFIPESPFGDPNPSGIAPLTPLOEQLOVFTETLEETGYLYIS 413
Qy 420 AWPDSLPLSVFONLQVIRGILHNGAYSILTLQGLIGISWGLRSLRELGSGLALIHNTH 479
Db 414 AWPDSLHDLVSFONLQVIRGILHNGAYSILTLQGLIGIRWGLRSLRELGSGLVLIHRNTH 473
Qy 480 LCFVHTVPWDQLFNPHQALLHTANRDECECVGGLACHQLCARGHCWGPPTQCVNCSQ 539
Db 474 LCFVHTVPWDQLFNPHQALLHSGNPSEEEGCKLDFACYPLCAHGCWGPPTQCVNCSH 533
Qy 540 FLRGQECVECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFPGPADQVCAHAHYKDP 599
Db 534 FLRGQECVKECRVWKGLPREVYNGKHCLPCHPECPQNSSTETCTGSEADQCTACPHYKDS 593
Qy 600 PFCVARGSPGVKPLSYMPIKFPDEBAGCOPCINCTHSCVDLDDKCPAQASPLTS 659
Db 594 PFCVARGSPGVKPLSYMPIWKYPDEGMQCPQPCINCTHSCVDLDERGCPAQASPLTS 653
Qy 660 IVSAVVGILLVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMR 719
Db 654 IIAVVGILLFLVIGVVVGLIKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQAMR 713
Qy 720 ILKETELRKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPGPKANKEILDEAYV 779
Db 714 ILKETELRKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPGPKANKEILDEAYV 773
Qy 780 MAGVSPVSVLLGLCLTSTVOLYTMOLMPYCILLDHVRENRGRIGSDLLNWCQIAKGM 839
Db 774 MAGLSPVSVRLGLCLTSTVOLYTMOLMPYCILLDHVRENRGRIGSDLLNWCQIAKGM 833
Qy 840 SYLEDVRLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIKWMALES 899
Db 834 SYLEDVRLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIKWMALES 893
Qy 900 ILRRRFTQSDVMSYGVVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM 959
Db 894 ILRRRFTQSDVMSYGVVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM 953
Qy 960 IMVKCWMIDSCRRFRELVSERFARMARDPQRFVVIQNEIDLGPASPLDSIFYRSLLEDDDD 1019
Db 954 IMVKCWMIDSCRRFRELVSERFARMARDPQRFVVIQNEIDLGPASPLDSIFYRSLLEDDDD 1013
Qy 1020 MGDVLDAEYLVPOQGFCDPPAPAGAGMVHHRSSSTRSGGGLDITLGLPEPSSEAPRS 1079
Db 1014 MGDVLDAEYLVPOQGFCDPPAPAGAGSTARRHRSSSTRSGGGLTLMGPEEPPRS 1073
Qy 1080 PLAPSEGAGSDVFDGLGMGAALKQLSLTHDPSPLQRYSDPTVPLPSETDGYVAPLTC 1139

Db 1074 PLAPSEGAGSDVFEGLGMGATKGPOSISPRDLSPLQRYSEDTPLPTSTDTGTVAPLAC 1133
Qy 1140 SPOPEYVQDFVDRPQPPSPREGPLPAARPAAGATLERAKTILSPKNGVVVDVFAFGGAVEN 1199
Db 1134 SPOPEYVQDFVDRPQPPSPREGPLPVRPAAGATLERKTLSPKNGVVVDVFTFGGAVEN 1193
Qy 1200 PEYLTPOGGAPOHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTTAENPEYLGLDV 1259
Db 1194 PEYLVPRGGSASQPH-PPALCPAFDNLVYWDQDPSERGSPPNTFEGTPTTAENPEYLGLDV 1252
Qy 1260 PV 1261
Db 1253 PV 1254
RESULT 3
TVRTNU
protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
C;Accession: A24562; A61204
R;Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A;Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.
A;Reference number: A24562; MUID:86118662; PMID:3945311
A;Accession: A24562
A;Molecule type: mRNA
A;Residues: 1-1260 <BAR>
A;Cross-references: EMBL:X03362; NID:956745; PIDN:CAA27059.1; PID:956746
R;Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen, S.
Carcinogenesis 12, 1975-1978, 1991
A;Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no m-
2-thiazolyl]formamide or N-methyl-N-nitrosourea
A;Reference number: A61204; MUID:92035293; PMID:1682063
A;Accession: A61204
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 637-663,'V',665-702 <MAS>
A;Note: authors translated the codon GCA for residue 25 as Val
C;Genetics:
A;Gene: neu
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphor-
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>
F;658-680/Domain: transmembrane #status predicted <TMN>
F;723-988/Domain: protein kinase homology <KIN>
F;731-739/Region: protein kinase ATP-binding motif
F;71,191,263,535,576,634/Binding site: carboxydrate (Asn) (covalent) #status predicted
F;691/Binding site: phosphate (Thr) (covalent) #status predicted
F;758/Active site: lys #status predicted
F;882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted
Query Match 86.2%; Score 5901; DB 1; Length 1260;
Best Local Similarity 86.2%; Pred. No. 2.7e-234;
Matches 1095; Conservative 48; Mismatches 105; Indels 22; Gaps 5;
Qy 1 MELAALCRWGLLLALLPPGAASSTVCTGTDMMKRLPASPTHLDMLRHLHYGCGQVVOGNL 60
Db 4 MELAAWCGWGLLLALLSPGASGTCTGTDMMKRLPASPTHLDIVRHLHYGCGQVVOGNL 63
Qy 61 ELTYLPTNASLSFLQDIQEQVFNFTVSFMLRVPKVSASHLEORLRIVRGTL 113
Db 64 ELTYVPANASLSFLQDIQEQV-----GYMLIAHNQVRVPL-----ORLRIVRGTL 110
Qy 114 FEDNYALAVLDNGDPLNNTTPTVT-GASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDT 172
Db 111 FEDKYALAVLDNRDPQDNVAASPTGRTPEGLRELQRLSLTEILKGGVLIQRNPOLCYQDM 170
Qy 173 ILWKDIFHKNNQLALTLIDNRSRACHPCSPMKGSRCWGESSEDCQSILRTVTCAGGCAR 232
Db 171 VLWKDVFRRKNQLAPVDIDNRSRACPPCAPACKDNHCWGPEDCQILTGTTCTSGCAR 230

QY 233 CKGPLEPTDCCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMNP 292
DB |||||
QY 231 CKGRLPTDCCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMNP 290
DB |||||
QY 293 GRYTFGASCVTACPNYLSTDVSGCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLG 352
DB |||||
QY 291 GRYTFGASCVTACPNYLSTDVSGCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLG 350
DB |||||
QY 353 MEHLREVRVANSIQEAFAGCKKIFGSLAFIPESFGDPDPAANTAPLOEQLOVETLEEI 412
DB |||||
QY 351 MEHLRGARAITSDNVQEDGCKKIFGSLAFIPESFGDPDPAANTAPLOEQLOVETLEEI 410
DB |||||
QY 413 TGYLYISAWPDSLPDLSVFNQLVTRGRLHNGAYSLTQGLGIFSWLGLRLSLRGSLA 472
DB |||||
QY 411 TGYLYISAWPDSLRDLSVFNQLVTRGRLHNGAYSLTQGLGIFSWLGLRLSLRGSLA 470
DB |||||
QY 473 LIHNTHLCFVHTVPWDQFLRNPHQALHTANRPEDE-CVGEGLACHQLCARGHCWGP 531
DB |||||
QY 471 LIHRNAHLCFVHTVPWDQFLRNPHQALHSGNRPEEDLCVSSGLVCNLSLCAHGHGWGP 530
DB |||||
QY 532 TQCVNCSOFLRGQECVECRVLQGLPREYVNAHCLPCHPCQCPQNGSVTCFGEADQCV 591
DB |||||
QY 531 TQCVNCSHFLRGQECVECRVWKGLPREYVSDRKLCPCHPCQCPQNGSVTCFGEADQCA 590
DB |||||
QY 592 ACAHYKDPFPCVAPCPGSKVDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGGPAE 651
DB |||||
QY 591 ACAHYKDSKCVAPCPGSKVDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGGPAE 650
DB |||||
QY 652 QRASPLTSIVSAVGVILLVWVGVFGLIKRRQKIRKYTMRLLQETELVEPLTPSGA 711
DB |||||
QY 651 QRASPVTFIATVEGVLLFLILVWVGVFGLIKRRQKIRKYTMRLLQETELVEPLTPSGA 710
DB |||||
QY 712 MPNAQMRILKETELRKVKVLSGAFGVYGIWIPDGENVKIPVAIKVLRENTSPKANK 771
DB |||||
QY 711 MPNAQMRILKETELRKVKVLSGAFGVYGIWIPDGENVKIPVAIKVLRENTSPKANK 770
DB |||||
QY 772 BILDEAYVMAGVSPYSRLIGICLTSTVOLVTLMPYGLCLDVRNHRGLSGQDLLNW 831
DB |||||
QY 771 BILDEAYVMAGVSPYSRLIGICLTSTVOLVTLMPYGLCLDVRNHRGLSGQDLLNW 830
DB |||||
QY 832 CMQIAGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP 891
DB |||||
QY 831 CVQIAGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP 890
DB |||||
QY 892 IKWMALESILRRRTHQSDVWSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPP 951
DB |||||
QY 891 IKWMALESILRRRTHQSDVWSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPP 950
DB |||||
QY 952 ICTIDVYIMVKWMIDSECPREFELVSEFSRWARDPQRFVWIONEDLGPASPLDSTFY 1011
DB |||||
QY 951 ICTIDVYIMVKWMIDSECPREFELVSEFSRWARDPQRFVWIONEDLGPASPLDSTFY 1010
DB |||||
QY 1012 RSLLEDDMDGLVDAEYLVLPQQGFCFDDPAPGAGGMVHRRHSRSTSGGGDLTLGLEP 1071
DB |||||
QY 1011 RSLLEDDMDGLVDAEYLVLPQQGFCFDDPAPGAGGMVHRRHSRSTSGGGDLTLGLEP 1070
DB |||||
QY 1072 SEEAAPRSPLAPSEGAGSDVFDGLGMCAAGLQSLPHTDPSPLQRYSEDTVPLPSETD 1131
DB |||||
QY 1071 SEEGPPRSPLAPSEGAGSDVFDGLGMCAAGLQSLPHTDPSPLQRYSEDTVPLPSETD 1130
DB |||||
QY 1132 GYVAPLTCSPQEVYNQDVPQPPSPREGPLPAARPAAGATLERAKTLSPCKNGVWVDVF 1191
DB |||||
QY 1131 GYVAPLTCSPQEVYNQDVPQPPSPREGPLPAARPAAGATLERAKTLSPCKNGVWVDVF 1190
DB |||||
QY 1192 AFGGAVENPEYLTQGGAAPOPHPPAFSPAFDNLTYWDQPPPPRGAPPSTFKGTPTAEN 1251
DB |||||
QY 1191 AFGGAVENPEYLVREGTASPPHSPAFSPAFDNLTYWDQNSSEGGPPSPNFGTPTAEN 1250
DB |||||
QY 1252 PEYLGLDVFPV 1261
DB |||||
QY 1251 PEYLGLDVFPV 1260
DB |||||

RESULT 4
GQHU
epidermal growth factor receptor precursor - human
N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C;Species: Homo sapiens (man)
C;Date: 15-Nov-1984 #sequence, revision 27-Nov-1985 #text change 11-Jun-1999
C;Accession: A00641; A25772; S30024; A38672; A00642; A05281; A05281; A33;
R;Ullrich, A.; Cousens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.; Ye,
P.H.
Nature 309, 418-425, 1984
A;Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression of
A;Reference number: A00641; MUID:84219729; PMID:6328312
A;Accession: A00641
A;Molecule type: mRNA
A;Residues: 1-1210 <ULL>
A;Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA25240.1; PID:g757924
A;Note: the authors translated the codon AAG for residue 540 as Asn
R;Ishii, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
A;Title: Characterization and sequence of the promoter region of the human epidermal growth
A;Reference number: A25772; MUID:85270438; PMID:2991899
A;Accession: A25772
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-29 <ISH>
A;Cross-references: GB:M11234; NID:g181981; PIDN:AA52370.1; PID:g553272
R;Haley, J.; Whittle, N.; Bennett, P.; Kitchington, D.; Ullrich, A.; Waterfield, M.
Oncogene Res. 1, 375-396, 1987
A;Title: The human EGF receptor gene: structure of the 110 kb locus and identification of
A;Reference number: S30024; MUID:88217333; PMID:3329716
A;Accession: S30024
A;Molecule type: DNA
A;Residues: 1-29 <HA2>
A;Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
R;Haley, J.D.; Waterfield, M.D.
J. Biol. Chem. 266, 1746-1753, 1991
A;Title: Contributory effects of de Novo transcription and premature transcript termination
A;Reference number: A38672; MUID:91107677; PMID:1988448
A;Accession: A38672
A;Molecule type: DNA
A;Residues: 1-29 <HAL>
A;Cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271
A;Experimental source: carcinoma cell line A431-7
R;Xu, Y.; Ishii, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.; Merl
Nature 309, 806-810, 1984
A;Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RNAs c
A;Reference number: A00642; MUID:84245835; PMID:6330563
A;Accession: A00642
A;Molecule type: mRNA
A;Residues: 150-187, 'KSUIQAV', 195, 'M', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-321,
, 798-799, 'TD', 802-811, 'R', 813-942 <XUY>
A;Experimental source: A431 human carcinoma cells, which have large numbers of EGF receptor
R;Lin, C.R.; Chen, W.S.; Krueger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verma, I.N.
Science 224, 843-848, 1984
A;Title: Expression cloning of human EGF receptor complementary DNA: gene amplification
A;Reference number: A43615; MUID:84196372; PMID:6326261
A;Accession: A43615
A;Molecule type: mRNA
A;Residues: 713-964 <LIN>
A;Experimental source: epidermoid carcinoma cell line A431
R;Simmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.
Biochem. Biophys. Res. Commun. 124, 125-132, 1984
A;Reference number: A23062; MUID:85046483; PMID:6093780
A;Accession: A23062
A;Molecule type: mRNA
A;Residues: 1028-1210 <SIM>
R;Weber, W.; Gill, G.N.; Speiss, J.
Science 224, 294-297, 1984
A;Reference number: A05281; MUID:84172183; PMID:6324343
A;Accession: A05281
A;Molecule type: protein
A;Residues: 25-30, 'S', 32-51, 454-467 <WEB>

R; Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.
 J. Biol. Chem. 260, 5205-5208, 1985
 A; Title: Identification of residues in the nucleotide binding site of the epidermal growth factor receptor. A60143; MUID:85182650; PMID:2985580
 A; Accession: A60143
 A; Molecule type: protein
 A; Residues: 740-744, 'X', 746-747 <RUS>
 R; Mroczkowski, B.; Mosig, G.; Cohen, S.
 Nature 309, 270-273, 1984
 A; Title: ATP-stimulated interaction between epidermal growth factor receptor and superoxide dismutase. A38023; MUID:84191554; PMID:6325948
 A; Contents: annotation; receptor activity
 A; Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
 R; Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.
 Cell 59, 33-43, 1989
 A; Title: Functional independence of the epidermal growth factor receptor from a domain x
 A; Reference number: A33331; MUID:9003233; PMID:2790960
 A; Contents: annotation; internalization signal
 C; Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor complex
 C; Genetics:
 A; Gene: GDB:EGFR
 A; Cross-references: GDB:120610; OMIM:131550
 A; Map position: 7p12.3-7p12.1
 C; Superfamily: epidermal growth factor receptor; protein kinase homology
 C; Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphotyrosine
 F; 1-24/Domain: signal sequence #status predicted <SIG>
 F; 25-1210/Product: EGF receptor #status predicted <MAT>
 F; 25-645/Domain: extracellular #status predicted <EXT>
 F; 730-600/Domain: EGF receptor extracellular domain repeat <EE2>
 F; 646-668/Domain: transmembrane #status predicted <TM>
 F; 669-1210/Domain: intracellular #status predicted <INT>
 F; 710-975/Domain: protein kinase homology <KIN>
 F; 718-726/Region: protein kinase ATP-binding motif
 F; 999-1046/Region: coated-pit mediated internalization signal
 F; 1047-1210/Region: inhibitory
 F; 128,175,352,413,444,528,603/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F; 745/Active site: Lys #status experimental

Query Match 45.2%; Score 3098; DB 1; Length 1210;
 Best Local Similarity 48.9%; Pred No. 1,1e-119;
 Matches 622; Conservative 180; Mismatches 356; Indels 114; Gaps 23;
 Qy 11 LLLALLPPGAA--STQVCTGTDMLRLPASPETHLDMLRLHYQGVVQVGNLEITYLPTN 68
 Db 14 LLAALCPASRALEEKVCQGSNKLTLQGTDFHFLSLQRMFNCEVVLGNLEITYVQRN 73
 Qy 69 ASLSFLQDIQVQ--FNNFTVFWLRVPKVSASHLEQRLIRIVRGTLFEDNALAVLNGD 127
 Db 74 YDLSFLKTIQVAGVYLIALNTVERIPL-----ENLQIRGNMYENSYALAVLSNYD 126
 Qy 128 PLNNTTPVTGASPGSLRELQRLSLTEILKGGVLIQRLNOLPCVQDTILWKDIFHKNNQAL 187
 Db 127 -----ANKTGLKELPWRNLQELIHGAVRFSNNPNCVSIQWRDIFSDFLSNM 176
 Qy 188 TLIDNTRSRACHPCSPMCKGRCRKGESSEDCQSLTRTVTCAGGCA-RCKGPLPTDCCHQC 246
 Db 177 SMDFQNLHSGCKQDPSCPNGSCWAGAEENCKLTKIICAQCCSRCKGKSPSCCHNQ 236
 Qy 247 AAGCTGPHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 306
 Db 237 AAGCTGPRSDCLVCRKFRDEATCKDCTPLMLYNPTTYQMDVNPPEGRYSGATCVKCKP 296
 Qy 307 YNLSLTDVGSCTLVCLPLHNQVETADGTPQCEKSKPCARCYGLGHEHLREVAVTASN 366
 Db 297 RNYVTDHSGSVCRAGADSYEM-EDGVRKCKCKGECGRKVCNGIGIGFEKDSUSINATN 355
 Qy 367 IQEFAGCKKIFGSLAFPESDGPASNTAPLQEQVFTLEBITGILYISAWPDSLP 426
 Db 356 IKHFKNCTSISSDLHILPVAFRGDSFTHPLDQELDLKTVKEITGFLLIQWPNRT 415
 Qy 427 DLSVFQNLQVIRGRILHNGAVSLTLQGLIGISWLGRLSLRELGSGLALIHNNHLCFVHT 486

Db 416 DLHAFENLEIRGRTKQHQGFSLAVSLNITSLGLRSLKLSIDGDIVIISGNKMLCYANTI 475
 Qy 487 PWDLFRNFHQALLHTANRREDECVGELACHOLCARGHCWGPGPTOCVNCNQFLOBOEC 546
 Db 476 NWKLLFGTSQKTKIISNRGENSKATGOVCHALCSPEGCGWPEPRDCVSCRNVSRREC 535
 Qy 547 VEBEVLQGLPREYVNAHCLPCHPEQOPQNGSVTCFPGPADQCACAHYKDPFPFCVARC 606
 Db 536 VDKCKLLEGPFRFVENSECICHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTC 595
 Qy 607 PSQVKPDLSPYMPKPFDEBGAQPCPINCSTHSCVDLDDKCCPAEQRASPLTSIVSAVG 666
 Db 596 PAGVGNENNTL-VNKYADAGVCHLCHPNCTYCTGPGLEGCTNGPKIP--SIATQWVG 652
 Qy 667 ---ILLVVVLGVVFGIILKRRQKIRKYTMRLQLQETELVEPLTPSGAMPNQAOIRLKE 723
 Db 653 ALLLLLVVAGIG---LFMRHRHVKRTLRLLQRELVEPLTPSGEAPNQLLRILKE 709
 Qy 724 TELRKVKVLGSGAFGVYKIGIWDGNGVKIPVAIKVLRNTPSKANKEILDEAYWAGV 783
 Db 710 TEFKKIKVLGSGAFGVYKGLWIPGEKVKIPVAIKELREATSPKANKEILDEAYWASV 769
 Qy 784 GSPVSVLLGLCTSTVOLTOLMPYCLLDHVRNRLGSLQDILLNMCQIAKMSYLE 843
 Db 770 DNPVHVCLLGLCTSTVQLITQLMPPFCGLDYYVREHKDNTGSGYLLNMCVQIAKMNYLE 829
 Qy 844 DVRLVHRDLAARNVVKSPNHVKITDFGLARLIDIDETEHADGGKVPKIKMALESILRR 903
 Db 830 DRLVHRDLAARNVVKTPQHVKITDFGLAKLGAEEKEYHAGGKVPKIKMALESILHR 889
 Qy 904 RFTHQSDVMSYGVTVVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDYVIMVK 963
 Db 890 IYTHQSDVMSYGVTVVWELMTFGSKPYDGIPIASISSILEKGERLPQPPICTIDYVIMVK 949
 Qy 964 CWMIDSECRPRRELVESEFSRMARDPQRFVVIQ-NEDLGASPLDSTFYRSLLEDDMDG 1022
 Db 950 CWMIDASRPFRELIEFESKARDPQRYLVIQGERMHLPSPTDSNFYALMDEEDMD 1009
 Qy 1023 LVDAEYLVFPQGFCCPDPAAGAGMVHRRHSRSGGDLTLGLPESEEEAPRPLA 1082
 Db 1010 VVDAEYLVFPQGF-----SSPSTSRPFL 1035
 Qy 1083 PSEGAGSDVDFDGLMGAAKGLQSLPDPSPLOQYSEDPTVPLPSET--DGVVAPLTC 1140
 Db 1036 SSLSATSN--NSTVACIDRNLQSCPIKEDSFLOQYSSDPTGALTEDSIDDTFL----- 1087
 Qy 1141 PQEYVNPQVDPQPSPREGPLPAAPAGATLERAKTLPKNGVVKVAFGGAENP 1200
 Db 1088 PVPEYINQ-SVPKRPAGSVQNVYHNQPLNP-----APSRDPHYQD--PHSTAVGNP 1136
 Qy 1201 EYL-TPQGAAPQHPHPPAFSPAFDNLVWDQ-----DP-----PERGAPPSTF 1243
 Db 1137 EYLVNTVQ-----PTCVNSTFDSPAHWAQKSHQISLDNPDYQDFFPKKAKPNGIF 1187
 Qy 1244 KGTPTAENPEYL 1255
 Db 1188 KGS-TAENAEYL 1198

RESULT 5

A53183
 epidermal growth factor receptor precursor - mouse
 C; Species: Mus musculus (house mouse)
 C; Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999
 C; Accession: A53183; A43818; S24942; A28941; S45325; I49643
 R; Luetteke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N.A.; I
 Genes Dev. 8, 399-413, 1994

A; Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor tyrosine kinase domain.
 A; Reference number: A53183; MUID:94170986; PMID:8125255
 A; Accession: A53183
 A; Molecule type: mRNA
 A; Residues: 1-1210 <LU>
 A; Cross-references: GB:U03425

RESULT 6
TVCHLV
epidermal growth factor receptor precursor - chicken
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C:Species: Gallus gallus (Chicken)
C>Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
C:Accession: A27720; A00643
R:Lax, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennstr
Mol. Cell. Biol. 8, 1970-1978, 1988
A:Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou
A:Reference number: A27720; MUID:82261272; PMID:3260329
A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: A00643
A:Molecule type: mRNA
A:Residues: 585-1223 <NIL>
A:Cross-references: GB:M10066
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
specific protein kinase
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F:31-654/Domain: extracellular #status predicted <EXT>
F:81-307/Domain: EGF receptor extracellular domain repeat <EE1>
F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F:655-677/Domain: transmembrane #status predicted <TM>
F:678-1223/Domain: intracellular #status predicted <INT>
F:719-984/Domain: protein kinase homology <KIN>
F:727-735/Region: protein kinase ATP-binding motif
F:136,202,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #s
F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:754/Active site: Lys #status predicted
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 44.6%; Score 3054.5; DB 1; Length 1223;
Best Local Similarity 48.0%; Pred. No. 6.6e-118;
Matches 625; Conservative 177; Mismatches 350; Indels 151; Gaps 28;
Qy 8 RWGULLALLPPGAA-----STQCTGDMKRLRPASPETHLMLRHLHYQCQVVOGNLE 61
Db 13 RGAAVLVLLGLGVALCSAVERKVCQGTNNKLTQLGHVEDHFTSLQRMNNCEVVLNLE 72
Qy 62 LTYLPTNASTLFDIOISQVQNNFTVSPWLRVPKVSASHLEQRLRIVRGTOIFEDNYALA 121
Db 73 ITYEHRNRLTLTKTIQEVA--GVVL---IALNMVDVIPLE-NIQIRGNVLYDNSFALA 126
Qy 122 VLDNGDPLNNTPVTGASPGGLRLQLRLSLTEILKGLGVLIQIRNPOLCYQDTILWKDIFHK 181
Db 127 VLSNVH-MNKQ-----GLRELPMKRLSEILNGGVKISNNPKLCNMOTVLDNDIIDT 177
Qy 182 NNQALTLID-TNRSRACHPCSPCKGSRCKGSESSDCQSLTRTVACAGCA-RCKGPLPT 239
Db 178 SRK-PLTVLDFASNLSLSSCPKHPNCTEDHCWGAGSQNCQTLTKVICAQCSGRCKGVPS 236
Qy 240 DCCHQCAAGCTGPKHSDCLACLNHNSGICELHCPALVTYNTDTPFSMPNPEGRYFGA 299
Db 237 DCCNHCQCAAGCTGPRSDCLACRFRRDATCKOTCPPLVLYNPITYQMDVNPESKYSFGA 296
Qy 300 SCVTACPNYLSLTDVSGCTLLCPHLNHOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREV 359
Db 297 TCVRECPHNVVTDHSGVRSNCNTDTEYV-EENGVRCKCKDGLGCKVNGIGIGELKGI 355
Qy 360 RAVTSANIQEPAGCKITFGSLAFIPESFGDGPASNTAPLOPEQLQVFTLEETGYLYIS 419

Db 356 LSNATNIDSFKNCTKINGDVSLPVAFLGDAFTKTLPLDPKLDVFTVKESGFLLIQ 415
Qy 420 AWPDSLPLSVFONLOVIRGILHNGAYSLLTQGLGISLWGLRSLRSLRGLSGLALIHENTH 479
Db 416 AWPDNATDLVAFENLEIRGRTKQHGQYSIAVNVNLIQSLGLSLKESISDGTAIMKNKN 475
Qy 480 LCFVHTVPMDQLFRNPHQALLTANPEDECVGEGGLACHOLCARGHCWGPGPTQCVNCSQ 539
Db 476 LCYADTNWME-SLFATQSQTKIIQNRKNNDCTADRHVCDPLCSDVGCWGPFPFCRCRF 535
Qy 540 FLQGECEVCEVRLQGLPREYVNAHCLPCHPECPONG---SVTCRGPADQCVACAHY 596
Db 536 FSRQKECVKCNILQCEPREFEREKLPCHECLVQNSTAYNTTCSGPGPDHCKMCAHF 595
Qy 597 KDPPFCVARCPGKVPDLSYMPITWKPDEGACQPCINCTHSCVDLDDGCPAEQASRP 656
Db 596 IDGPHCVKACPAVGLGENDTL-VKXADANAVCOLCHPNCTRGCKGFLGSCP---NGSK 651
Qy 657 LTSIVSAVV-GILLVVVLGVVFGILIKRRQKIRKYTMRLLOETELVEPLTPSGAMPNQ 715
Db 652 TPSIAGVGGLLCLLVVGLIGLYLRRR-HIVKRTLRLLQERELVEPLTPSGEAPNQ 710
Qy 716 AQMRILKETELRKVKVLGSGAFGTVYKGIWIPGENVKIPIVAIKVLRENTSPKANKILD 775
Db 711 AHLRIILKETEFKKVKVGLSGAFGTVYKGLWIPGEKVKIPVAIKELREATSPKANKILD 770
Qy 776 EAYMAGVGSPPYRLLIGLICLTSTVOLATOLMPYVGLLDHVRNRLGSGDILLNMCQI 835
Db 771 EAYMASVDNPHVCRLLIGLICLTSTVQITOLMPYVGLLDYIREHKONIGSOYLLNMCVQI 830
Qy 836 AKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKWM 895
Db 831 AKGMYLEERLVRDLAARNVLVKTPQHVKITDFGLAKLLGADEKEYHAEGGKVPKWM 890
Qy 896 ALESILRRRTHOSDVWSYGVTVWELMTFGAKYDGIPIAREIPDLLEKGBRLPOPPCTI 955
Db 891 ALESILHRIYTHQSDVMSYGVTVWELMTFGSKYDGIPIASEISSVLEKGBRLPOPPCTI 950
Qy 956 DVYIMVWKMIWIDSECRPRELVSFSSRMARDPQRFVVIQ-NEDLGPASPLDSTFVRS 1014
Db 951 DVYIMVWKMIWIDSECRPRELVSFSSRMARDPQRFVVIQ-NEDLGPASPLDSTFVRS 1010
Qy 1015 LEDDDMGDLVDABEYLVPQOQFFCPDPAPGAGGMVHRRHSSSTSRSGGDLTLGLEPSEE 1074
Db 1011 MEEDMEDIVDAEYLVPHQGF-----NSPST----- 1038
Qy 1075 EAPRSL-----APSEGAGSDVFDGLGMAAGKGLSLPHDPSPLORYSEDPVLPSE 1129
Db 1039 --SRTPLLSLSATSNNATNCID-----RNGOCHPVREDSFVORYSSDPTGNFLEE 1088
Qy 1130 T--DGVAVPLTCSPOEYVNPQDVRPOPSPREGPLPAARPAGATLERAKTLSPGKNGV 1187
Db 1089 SIDDGEU-----PAPEYVNOQ--LMPKKPS-----TAMVQOIY 1119
Qy 1188 KDVF-----AFGAVENPEYLTPOGGAAPQHPPPPAPFPAFDNLYWDQ- 1231
Db 1120 NNISLTAISKLPMDSRVQNSHSTAVDNPEYL-----NTNOSPPLAKTVFESSPYWIOS 1171
Qy 1232 -----DPPE-----RGAPSTFKGTPTAENPEYLGLDVP 1260
Db 1172 GNHQINLDNPDYQDDFLPNETKPNGLLKVPAAENPEYLRVAAP 1214

RESULT 7

A47253

epidermal growth factor receptor, HER4 - human

C:Species: Homo sapiens (man)

C>Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999

C:Accession: A47253

R:Florman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; Ne

Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993

A:Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal

Db 514 PQQCLSCRNYSREGVCTVCHNPLQGEPRFVHEAQCFCHCEPLDMEGTSTYNSGGSDAC 573
Qy 591 VACAHYKDPFPCVACPSGVKPDLSYPIWKFDPDEGACQCPINCTHSC--VLDLDDKGC 648
Db 574 ARCAHFRDGPCHVNSCPHILG--AKGPIKYVPAQNECRFCHENTQCGNGPELQDCIG 631
Qy 649 PAEQBASPLTSIVSAVGVILLVVLGVVFGVILIKRQOKIR-KYTMRLLOETELVEPLT 707
Db 632 QAEVLMKSPHIVIAVTVG--LAVILMLGGSFYWRGRRIQKRAMRYLERGESIEPLD 689
Qy 708 PSGAMPNOAQRILKETELRKVKVLGSGAFGVYKGIWIPDGNVKIIPVAIKVLRNTSP 767
Db 690 PS-EKANKVLARIKETELRKVKVLGSGVFGVHKGIWIPGESIKIPVCIKVIEDKSGR 748
Qy 768 KANKEILDEAYVMAGVGSPPYVSRLLGICLTSTVOLVQLMYPYGCILLDHVRENRLGSQD 827
Db 749 QSFQAVTDHMLAVGSLDAHVIRLLGLCPGSSQLVQVLYPLGSLLDHVQKHRETLGPQL 808
Qy 828 LLNWCMIQAKMSYLEVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADG 887
Db 809 LLNMGVQIAKMGYYLEESHSVHRDLAARNVLMKSPSQVQVADFGVADLLPPDDKQLLHSE 868
Qy 888 GKVPILKMALESILRRRTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERL 947
Db 869 AKTPIKMALESILHFKTHOSDVMSYGVTVWELMTFGAEFYAGLRLAEIPDLLEKGERL 928
Qy 948 PQPPICTIDVYMWKMWIDSECRPRFELVSEFSRMARDPQRFVVTIONEDLGPASPLD 1007
Db 929 AQPQICTIDVYMWKMWIDENIRPTPKELANEFTRWARDPPVLYVIKRAS-GPGTP-- 985
Qy 1008 STFRSLLEDMDGDLVDAEYLVPOQGFPCDPAPGAGGVVHHRHRSSTSTRSGGGLTL 1067
Db 996 PAAPFSLVTTKEL--QEALEPEL-----DLDL 1011
Qy 1068 GLEPSEE-----EAPSPAPSEG-----AGSDVFDGLGMG 1099
Db 1012 DLEAEGLATLSGALSALPTGTTTRPGSOLSPSSGYPMQSSLGEACLDSAVLGG 1071
Qy 1100 AAKGLQLTHDPSPLOQRYSEDPTVLPSETDGYV---APL-----TC-----SPQ 1143
Db 1072 REQFSRPISLH-PIPRGR-----PASESEGHVTGSEAELOEKVSVCRSRSRSPRP 1123
Qy 1144 E-----YVNPQVVRQPPSPREGP-----LPAARPAGATLERAKTLSP-GKNGV 1187
Db 1124 RGSAYHSQRHSLTPTVPLPSPGLEEDNGYVMPDTHLRGASSRREGTLSSVGLSSVL 1183
Qy 1188 -----KDVFAFGGAVENPEYLTQGGGAPOPHP 1216
Db 1184 GTEEEDE-----EEYEMNRKRKRGSP-PRPP 1209

RESULT 11
TVFVLV
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
N:Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C:Species: avian leukosis virus, ALV
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C:Accession: B00643; A00643
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: B00643
A:Molecule type: mRNA
A:Residues: 1-698 <NTL>
A:Cross-references: GB:M10066; GB:M13881; NID:g211749; PIDN:AAA48763.1; PID:g211750
A:Note: in Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gal
C:Comment: This protein is synthesized as a gag-env-erbB protein.
C:Genetics:
A:Gene: gag-env-erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
F:1-6/Product: gag protein (fragment) #status predicted <GAG>

F:7-59/Product: env protein (fragment) #status predicted <ENV>
F:60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F:194-459/Domain: protein-kinase homology <KIN>
F:202-210/Region: protein kinase ATP-binding motif
F:229/Active site: Lys #status predicted

Query Match 25.8%; Score 1766.5; DB 1; Length 698;
Best Local Similarity 52.2%; Pred. No. 1.8e-65;
Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18;

Qy 584 GPEADQCVACAHYKDPFPCVACPSGVKPDLSYPIWKFDPDEGACQCPINCTHSCVDL 643
Db 60 GP--DHCKCAHFIDGPHCVKACPAAGVLGENDTL-VMKYADANAVCOLCHPNCNTRGCKGP 116
Qy 644 DDGCPAEQAEASPLTSIVSAV-GILLVVLGVVFGVILIKRQOKIRKYTMRLLOETEL 702
Db 117 GLECP---NGSKTPSIAAGVVGGLLCLLVVGLGIGLYLRRR-HIVKRTLRLLOEREL 172
Qy 703 VEPLTPSGAVPNOAQRILKETELRKVKVLGSGAFGVYKGIWIPDGNVKIPVAIKVLR 762
Db 173 VEPLTPSGEAPNOAHLRIKETEFKVKVLGSGAFGVYKGLWIPEGEKVKIPVAIKELR 232
Qy 763 ENTSPKANKEILDEAYVMAGVGSPPYVSRLLGICLTSTVOLVQLMYPYGCILLDHVRENGR 822
Db 233 EATSPKANKEILDEAYVMAGVGSPPYVSRLLGICLTSTVOLVQLMYPYGCILLDIYREHON 292
Qy 823 LGSODLLNWCMIQAKMSYLEVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 882
Db 293 IGSQYLLNWCMIQAKMSYLEVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 352
Qy 883 YHADGGKVPILKMALESILRRRTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLE 942
Db 353 YHAEKGKVPILKMALESILHRIYTHOSDVMSYGVTVWELMTFGSKPYDGIIPAREIPDLLE 412
Qy 943 KGERLPOPPICTIDVYMWKMWIDSECRPRFELVSEFSRMARDPQRFVVIQ-NEDLG 1001
Db 413 KGERLPOPPICTIDVYMWKMWIDSECRPRFELVSEFSRMARDPQRFVVIQ-NEDLG 472
Qy 1002 PASPLDSTFRSLLEDMDGDLVDAEYLVPOQGFPCDPAPGAGGVVHHRHRSSTSTRSG 1061
Db 473 LPSTDSKFTLMEEDMEDIVDAEYLVPHOGFF-----NSPST--- 513
Qy 1062 GGDLTGLLEPSEBEAPRSL-----APSEAGSDSDVFDGLGMGAQGLSLPHTDPSPLQ 1116
Db 514 -----SRTPLLSLSATSNNSSATNCID-----RNGQGHVPVREDSFVQ 550
Qy 1117 RYSEDDPTVLPSEET--DGYVAPLTCSPQPEYVNPQVVRPQPPSPREGPLPAARAGATLE 1174
Db 551 RYSSDPTGNFLBSIDGDL-----PAPEYVNO--LMPKKPS----- 585
Qy 1175 RAKTLSPGKGVVVKDVF-----AFGGAVENPEYLTQGGGAAPQPHPPAF 1219
Db 586 -----TAMVQNIYNNISLTAISKLPMSRQNSHSTAVDNPEYL-----NTNQSLA 633
Qy 1220 SPADFNLVYDQ-----DPPE-----RGAPPSTFGKTPAENPEYLGLDVP 1260
Db 634 KTVFESSPYWQSGNHQINLNDPDYQDQDFLPNETKPNGLLKVPAAENPEYLRVAAP 689

RESULT 12

TVYUH

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)

C:Species: avian erythroblastosis virus

C:Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999

C:Accession: A00644; A38022

R:Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.

Cell 35, 71-78, 1983

A:Title: The erbB gene of avian erythroblastosis virus is a member of the src gene family

A:Reference number: A00644; MUID:84026539; PMID:6313229

A:Accession: A00644

A:Molecule type: DNA

A:Residues: 1-604 <YAM>

A:Cross-references: GB:K01216; NID:g209676; PIDN:AAA42400.1; PID:g209678

Db 502 QMHLRLWLXYIQVINSSTQDSNEHQLTDACYSPSVPTSLTIERARYAIQISAGLAMELE 561
Qy 511 -----VGEGLA-----CHQLCARGHC 526
Db 562 QITARSASMRHSKTLPAEGRQVPRWVFLGVCASARAGIAEPLAGRAVCRKCHPCELCTN 621
Qy 527 WGPQPTQCVNCSQFLRGQECVECRVLOGLPREYV---NARHCLPCHPECPQNGSVTCF 583
Db 622 YGYHEQVCSKTHYKRRQCECTEC-----PADHYTDEQRECFORHPEC---NG---CT 669
Qy 584 GPEADQCVACAHYK-----DPPF-----CVARCPGVK-PDLSYMPIWKFPDEEGAC 629
Db 670 GPGADDCKSCRFNFKLFDANETGPVYNSTFMNCTSKCPLMRHVNYYTAIGPY-----C 723
Qy 630 QPCPINCTHSCVDLDDKCPAEQASPLTSIVSAVVGILLVVLGVFGI-LIKRROOKI 688
Db 724 AASPFRSKITANLD-----VNNIFITGAVLPTCILCVVYIICRQOKA 770
Qy 689 RKYT--MRLLOETELVEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWI 746
Db 771 KKETVKTMTALSGREDSEPLRPSNIGANLCKLRIVKDAELRKGVLGMAFGRVYKGVWV 830
Qy 747 PDGENVKIPVAIKVRENTSPKANKEILDEAYVMAGVSPVSRLLGICLTSTVQLVTOL 806
Db 831 PEGENVKIPVAIKELLKSTGAESSEFLREAYIMASEHVNLLKLLAVCMSSQWMLITOL 890
Qy 807 MPYGLLDHVRNRRGLSGQDLLNWCQIAKMSYLEVDVLRVHRDLAARNVLVK---SPN 863
Db 891 MPLGLLDYVNRNDRKIGSKALLNWSQIAKMSYLEEKLVRVHRDLAARNVLVRLLAGED 950
Qy 864 HVKITDGLARLLDIDETEHADGGKVPKIKMALESILRRRFTHQSDVMSYGVTVWELMT 923
Db 951 H---DFGLAKLLSSDSNEYKAAGKMPKWLALCIRNRVFTSKSDVWAFGVTIWELLT 1006
Qy 924 FGAKPYDGPARETDLLEKERLPPOPPICTIDVYIMVKCWMIDSECRPRFRELVSF 983
Db 1007 FGQRPHENIPAKDIPDLIEVGLKLEQPEICSLDYCTLLSCWHLDAAMRPTFKQLTVFA 1066
Qy 984 RMARDPQRFVVIQNEGLG--PASPLDSTFYRSLLEDD---DMGLVDAAEYLVFQOGFFC 1038
Db 1067 EFADPGRYLAIGDKKTRLPA-----YTSQEKDLIRKLAPTTDGEAIAKPDYLYQ 1119
Qy 1039 PDPAAGAGVHHRHRSSTRSKSGGDLTLGLEPSEERAP-----RSLAPSEAGAGSDVF 1092
Db 1120 PKAALGPS-----HRTDCT-----DEMPKLNRYCKDPSNKNKSTGDDER 1158
Qy 1093 DG---DLGMGAAGLQSLPHDPSLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYVNOP 1149
Db 1159 DSSAREVGVGNLR-----LDLPVDEDDYIMP--TCQPGENNNNM 1196
Qy 1150 DVPQPPSPREGPLPAARPAATLERAKTLSPGKNGVVKDVFAFGGAVENPEYL---TP 1205
Db 1197 N-----NPNQNNMAAGVAAGYM-----DLIGVPVSDNPEYLLNAQTL 1235
Qy 1206 QGGAAPQPH-----PPAFSP-AFDNLYYWD 1230
Db 1236 GVGESPIPTQIGIPVGGPGTMEVKVMPGSEPTSSDHEYND 1279

RESULT 14

S35745

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus

C:Species: avian erythroblastosis virus

C:Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997

C:Accession: S35745

R:Vennstrom, B.

submitted to the EMBL Data Library, March 1993

A:Reference number: S35743

A:Accession: S35745

A:Molecule type: DNA

A:Residues: 1-544 <VEN>

A:Cross-references: EMBL:X12707

C:Genetics:

A:Gene: erbB

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: Atp; oncogene; phosphotransferase; transforming protein; tyrosine-specific p

F;135-400/Domain: protein kinase homology <KIN>

F;143-151/Region: protein kinase ATP-binding motif

F;170/Active site: Lys #status predicted

Query Match 24.1%; Score 1647; DB 2; Length 544;
Best Local Similarity 54.9%; Pred. No. 1.1e-60;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

Qy 584 GPEADQCVACAHYKDPFPCVARGSPGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDL 643
Db 1 GP--DHCMKAHFIDGPHCVKCAPAGVLGENDTL-VMKYADANAVACQLCHPNCNTRGCKGP 57
Qy 644 DDKGCPEAQASPLTSIVSAV-VGILLVVLGVFGI-LIKRROOKIKRQKIMRRLLOETEL 702
Db 58 GLEGCP---NGSKTPSIAAGVVGGLCLVVGIGIYLRRR-HIVRKRITLRRLLQREEL 113
Qy 703 VEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDCGENVKIPVAIKVLR 762
Db 114 VEPLTPSGEAPNQAHLRLKETEFKVKVLGFGAFGTVYKGLWIPGEKVTIPVAIKELR 173
Qy 763 ENTSPKANKEILDEAYVMAGVSPVSRLLGICLTSTVQLVTQMPYGCLLDHYVRENRR 822
Db 174 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDN 233
Qy 823 LGSQDLLNWCQIAKMSYLEVDVLRVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 882
Db 234 IGSQYLLNWCQVIAKGMNLYLEERHVMHRDLAARNVLKTPHQVKITDFGLAKQLGADEKE 293
Qy 883 YHADGGKVPKIKMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI-PAREIPDLLE 942
Db 294 YHAEKGKVPKIKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPDGIPASISSVLE 353
Qy 943 KGERLPPOPPICTIDVYIMVKCWMIDSECRPRFRELVSFMRMARDPQRFVVIQ-NBDLG 1001
Db 354 KGERLPPOPPICTIDVYIMVKCWMSDASRPKRELIATAEFSKWARDPPRVLIQCDERMH 413
Qy 1002 PASPLDSTFYRSLLEDDMGDLVDAAEYLVFQOGFFCDDPAPGAGGVHHRHRSSTRS 1061
Db 414 LPSPTDSKFRYTLMEEDMEDIVDADEYLVPHQGF-----NSPST--- 454
Qy 1062 GGDLTLGLEPSEERAPSL-----APSEGAGSDVFDGDLGMGAAGLQSLPHDPSLQ 1116
Db 455 -----SRTPLLSLSATSNNATNCIDRNG-----H----- 481
Qy 1117 RYSEDPTVPLPSETDGVVAPLTCSPQPEYVNOPDVRPQPPSPREGPLPAARPAAT-LE 1175
Db 482 -----FVREDGFL-----PAPEYVQ--LMPKPSAMVQVQIYNIISLTAISK 523
Qy 1176 AKTLPQKNGVVKDVFAFGGAVENPEYL 1203
Db 524 LPIDSRYQN-----SHSTAVDNPEYL 544

RESULT 15

S00727

kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis virus (

C:Species: avian erythroblastosis virus

C:Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997

C:Accession: S00727

R:Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.

Oncogene Res. 1, 265-278, 1987

A:Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mutant

A:Reference number: S00727; MUID:88217326; PMID:2897102

A:Accession: S00727

A:Molecule type: DNA

A:Residues: 1-545 <SCO>

A:Cross-references: EMBL:X06943

C:Genetics:

A:Gene: erbB

Search completed: July 22, 2003, 09:08:10
Job time : 31.0157 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2793 Seconds
(without alignments)
5088.033 Million cell updates/sec

Title: SEQ4-59-73-14

Perfect score: 6847

Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDPV 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	6707	98.0	1255	1 ERB2 HUMAN	P04626 homo sapien
2	5907	86.3	1257	1 ERB2 RAT	P06494 rattus norv
3	5901.5	86.2	1254	1 ERB2 MESAU	Q00553 mesocricetu
4	3096	45.2	1210	1 EGFR HUMAN	P00533 homo sapien
5	3075	44.9	1210	1 EGFR MOUSE	Q01279 mus musculu
6	2930.5	42.8	1308	1 ERB4 HUMAN	Q15303 homo sapien
7	2911	42.5	1308	1 ERB4 RAT	Q62956 rattus norv
8	2649.5	38.7	1167	1 XMRK_XIPWA	P13388 xiphophorus
9	2382.5	34.8	1342	1 ERB3 HUMAN	P21860 homo sapien
10	2316.5	33.8	1339	1 ERB3 RAT	Q62799 rattus norv
11	1919	28.0	1426	1 EGFR DROME	P04412 drosophila
12	1749.5	25.6	634	1 ERBB ALV	P00534 avian leuko
13	1703	24.9	604	1 ERBB AVIER	P00535 avian eryth
14	1630	23.8	540	1 ERBB AVIEU	P11273 avian eryth
15	1555	22.7	703	1 EGFR CHICK	P13387 gallus gall
16	1274	18.6	1323	1 LT23 CABEL	P24348 caenorhabdi
17	1142.5	16.7	245	1 ERB2 MOUSE	P70424 mus musculu
18	736	10.7	1363	1 ILPR BRALA	O02466 branchiosto
19	718	10.5	1382	1 INSR HUMAN	P06213 homo sapien
20	713	10.4	1383	1 INSR RAT	P15127 rattus norv
21	712.5	10.4	1372	1 INSR MOUSE	P15208 mus musculu
22	706	10.3	1300	1 IRR MOUSE	Q9wt14 mus musculu
23	698	10.2	1297	1 IRR HUMAN	P14616 homo sapien
24	692.5	10.1	1300	1 IRR_CAVPO	P14617 cavia porce
25	691	10.1	1607	1 MIPR LYNST	Q25410 lymnaea sta
26	670	9.8	1477	1 HTK7 HYDAT	Q25197 hydra atten
27	669	9.8	1367	1 IG1R HUMAN	P08069 homo sapien
28	659	9.6	1373	1 IG1R MOUSE	Q60751 mus musculu
29	655.5	9.6	1370	1 IG1R RAT	P24062 rattus norv
30	623	9.1	1390	1 INSR_AEDAE	Q93105 aedes aegyp
31	613	9.0	987	1 EPB4 HUMAN	P54760 homo sapien
32	613	9.0	2146	1 INSR_DROME	P09208 drosophila
33	595	8.7	987	1 EPB4_MOUSE	P54761 mus musculu

RESULT 1							
ERB2_HUMAN	34	591.5	8.6	984	1	EPB1_CHICK	Q07494 gallus gall
AC	35	589.5	8.6	977	1	EPB2_MOUSE	Q03145 mus musculu
DT	36	588	8.6	1114	1	RET_HUMAN	P07949 homo sapien
DT	37	585	8.5	984	1	EPB1_RAT	P09759 rattus norv
38	584.5	8.5	976	1	EPB2_HUMAN	P29317 homo sapien	
39	580	8.5	984	1	EPB1_HUMAN	P54762 homo sapien	
40	579	8.5	902	1	EPBB_XENLA	Q91736 xenopus lae	
41	578.5	8.4	985	1	EPBA_XENLA	Q91571 xenopus lae	
42	573.5	8.4	1053	1	PAK1_CHICK	Q00944 gallus gall	
43	569	8.3	1068	1	PAK1_XENLA	Q91738 xenopus lae	
44	567	8.3	757	1	HT16_HYDAT	P53356 hydra atten	
45	563	8.2	1052	1	PAK1_MOUSE	P34152 mus musculu	
ALIGNMENTS							
STANDARD; PRT; 1255 AA.							
DT 13-AUG-1987 (Rel. 05, Created)							
DT 13-AUG-1987 (Rel. 05, Last sequence update)							
DT 15-JUN-2002 (Rel. 41, Last annotation update)							
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)							
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell							
DE surface receptor HER2) (MLN 19).							
GN ERBB2 OR HER2 OR NGL OR NEU.							
OS Homo sapiens (Human).							
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.							
OX NCBI_TaxID=9606;							
RN [1]							
RP SEQUENCE FROM N.A.							
RX MEDLINE=86118663; PubMed=3003577;							
RA Yanamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,							
RA Saito T., Toyoshima K.;							
RT "Similarity of protein encoded by the human c-erb-B-2 gene to							
RT epidermal growth factor receptor.;"							
RL Nature 319:230-234(1986).							
RN [2]							
RP SEQUENCE FROM N.A.							
RX MEDLINE=86070181; PubMed=2999974;							
RA Coussens L., Yang-Feng T.H., Liao Y.C., Chen E., Gray A.,							
RA McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,							
RA Francke U., Levinson A., Ullrich A.;							
RT "Tyrosine kinase receptor with extensive homology to EGF receptor							
RT shares chromosomal location with neu oncogene.;"							
RL Science 230:1132-1139(1985).							
RN [3]							
RP SEQUENCE OF 737-1031 FROM N.A.							
RX MEDLINE=86016729; PubMed=2995967;							
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.;							
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the							
RT c-erbB-1/epidermal growth factor-receptor gene and is amplified in a							
RT human salivary gland adenocarcinoma.;"							
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).							
RN [4]							
RP VARIANTS VAL-654 AND VAL-655.							
RX MEDLINE=93194196; PubMed=8095488;							
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;							
RT "Characterization of a new allele of the human ERBB2 gene by allele-							
RT specific competition hybridization.;"							
RL Genomics 15:426-429(1993).							
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,							
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A							
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-							
CC ALPHA AND AMPHIREGULIN.							
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyro							
CC tyrosine phosphate.							
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS							
CC (POTENTIAL).							
CC -1- SUBCELLULAR LOCATION: Type I membrane pr							


```
QY 721 LKETLRKVKVLGSGAFGVYKGIWI PDGENVKIPVAIKVLRNTPSKANKEIIDEAYVM 780
D 715 LKETLRKVKVLGSGAFGVYKGIWI PDGENVKIPVAIKVLRNTPSKANKEIIDEAYVM 774
QY 781 AGVGSPPVSRLLGICLSTVOLVTQLMPYGLLDHVRNRCGLSQDLLNWCMIAGKMS 840
D 775 AGVGSPPVSRLLGICLSTVOLVTQLMPYGLLDHVRNRCGLSQDLLNWCMIAGKMS 834
QY 841 YLEDRVLRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMWALESI 900
D 835 YLEDRVLRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMWALESI 894
QY 901 LRRRFTHQSDVMSYGVTVWELMTGAKPYDGIPIAREIPDLEKGERLPQPPICITDVYMI 960
D 895 LRRRFTHQSDVMSYGVTVWELMTGAKPYDGIPIAREIPDLEKGERLPQPPICITDVYMI 954
QY 961 MVKCMWMDISCRPRFRELVSFSEWARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1020
D 955 MVKCMWMDISCRPRFRELVSFSEWARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1014
QY 1021 GDLVDAEYLVPQOGFFCPDPAPGAGGVHHRHRSSTRSGGDLTLGLEPSEEEAPRSP 1080
D 1015 GDLVDAEYLVPQOGFFCPDPAPGAGGVHHRHRSSTRSGGDLTLGLEPSEEEAPRSP 1074
QY 1081 LAPSEGAGSDVFDGDLGWAAGKGLQSLPTHDPSPQLQRYSEDPTVPLPSETDGYVAPLTCS 1140
D 1075 LAPSEGAGSDVFDGDLGWAAGKGLQSLPTHDPSPQLQRYSEDPTVPLPSETDGYVAPLTCS 1134
QY 1141 POPEYVNPQDVRPOPSPREGPLPAARPAAGATLERAKTLPKNGVVKDVPFAFGGAVENP 1200
D 1135 POPEYVNPQDVRPOPSPREGPLPAARPAAGATLERAKTLPKNGVVKDVPFAFGGAVENP 1194
QY 1201 EYLTPQCGAAPQHPHPPAFSPAFNLVYWDQDPERGAPSTPKGTPTAENPEYVLGLDVP 1260
D 1195 EYLTPQCGAAPQHPHPPAFSPAFNLVYWDQDPERGAPSTPKGTPTAENPEYVLGLDVP 1254
QY 1261 V 1261
D 1255 V 1255

RESULT 2
ERB2 RAT STANDARD; PRT; 1257 AA.
AC P06494;
DT 01-JAN-1988 (Rel. 06, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (update)
DE (p185erbB2). (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor
DE receptor-related protein).
DE ERBB2 OR NEU.
GN Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuroblastoma;
RX MEDLINE=86118662; PubMed=3945311;
RA Bargmann C.I., Hung M.-C., Weinberg R.A.;
RT "The neu oncogene encodes an epidermal growth factor receptor-related
RT protein.";
RL Nature 319:226-230 (1986).
RN [2]
RP SEQUENCE OF 852-905 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704 (1991).
```

```

RN [3]
RP STRUCTURE BY NMR OF 650-668.
RX MEDLINE=92155181; PubMed=1346763;
RA Gullick W.J., Bottomley A.C., Loftis F.J., Doak D.G., Mulvey D.,
RA Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
RT "Three dimensional structure of the transmembrane region of the proto-
RT oncogenic and oncogenic forms of the neu protein.";
RL EMBO J. 11:43-48 (1992).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PFM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03362; CAA27059.1; ALT_INIT.
DR PIR; A24562; TVRTNU.
DR HSPSP; P11362; IFGK.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00769; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; Fu; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Proto-oncogene; Disease mutation.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1257 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT DOMAIN 22 654 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 655 677 POTENTIAL.
FT DOMAIN 678 1257 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 159 369 CYS-RICH.
FT DOMAIN 473 646 CYS-RICH.
FT DOMAIN 722 989 PROTEIN KINASE.
FT NP_BIND 728 736 ATP (BY SIMILARITY).
FT BINDING 755 755 ATP (BY SIMILARITY).
FT ACT_SITE 847 847 BY SIMILARITY.
FT DISULFID 196 205 BY SIMILARITY.
FT DISULFID 200 213 BY SIMILARITY.
FT DISULFID 221 228 BY SIMILARITY.
FT DISULFID 225 236 BY SIMILARITY.
FT DISULFID 237 245 BY SIMILARITY.
FT DISULFID 241 253 BY SIMILARITY.
FT DISULFID 256 265 BY SIMILARITY.
FT DISULFID 269 296 BY SIMILARITY.
FT DISULFID 300 312 BY SIMILARITY.
FT DISULFID 316 332 BY SIMILARITY.
FT DISULFID 335 339 BY SIMILARITY.
FT DISULFID 513 522 BY SIMILARITY.
```


Db 341 SLMSAQTVDSNDIKFNCTKINGNLFLVLTGIGHDPYNALEADPEKLVNVRTVREITG 400
Qy 415 YLYISAMPDLPDLVSFQNLQVIRGRILHNGAYSITLQGLGISWGLRSLRSLGSLALI 474
Db 401 FLNIQSPNNMTDFSVSNLVTIGRVLVSLGLSLLILKQQTSLQFQSLKEISAGNIYI 460
Qy 475 HHNTHLCFVHTVPNDOLFNRPHQALLHTANRPEDECVEGSLACHOLCARGHCWPGPTQC 534
Db 461 TDNSNLCCYHTINWTLFTSTINQIRIVDRNKAENCTAEGMVCNHLCSDDCGWPGPDQC 520
Qy 535 VNCQFRLRGQCBECRVLQGLPREYVNAHCLPCHPECQ-ONGSVTCFGEADQCVCAC 593
Db 521 LSCRRFGRICIESCNLYDGEFREFNGSICVCEDCQCEKMEGLTCHGPGPDNCTKC 580
Qy 594 AHYKDPFCVACPSGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGC----- 648
Db 581 SHFKDGENCBEKCPDGLQGANF--IPKYADPDRECHPCNCTQGCNGPTSHDCIYYPW 638
Qy 649 -----PAEQBASPLTSIVSAVV--GILLVVVLGVVFGILIKRQOKTRKYTMRLLQETE 701
Db 639 TGHSTLPQAR--TFL--IAAGVIGGLFVLVGLTFVAVYVRKSIK-KGRALRFL-ETE 693
Qy 702 LVEPLTSGAMPNOAMRILKETELRKVKVLGSGAFGVYKGIWIPGENVKIPVAIKVL 761
Db 694 LVEPLTSGTAPNOAQRIKETELRKVKVLGSGAFGVYKGIWIPGETVKIPVAIKIL 753
Qy 762 RENTSPLANKEILDEAYVMAGVGSPPVSRLLIGLICLTSTVLQVLTQMPYGCILLDHVRENRG 821
Db 754 NETTGKANVEFMEALIMASMDPHLVRLLGVCLSPITQVLTQMPHGCILLEYVHEKD 813
Qy 822 RLGSQDILNMCWTAKGMSYLEDEVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDET 881
Db 814 NIGSOLLNMCVQITAKGMVLEERLVRDLAARNVLKSPNHVKITDFGLARLLGDEK 873
Qy 882 EYHADGKVPITKMWALSIILRRRTHQSDVMSYGVTVWELMTCGAKPDGIPAREIPDLL 941
Db 874 EYNADGKMPKMWALSIILRRRTHQSDVMSYGVTVWELMTCGAKPDGIPAREIPDLL 933
Qy 942 EKERLPQPPCTIDVTYIMVWKCMIDSECRPREFRELVSFERNMARDPQRFVITQNRD-L 1000
Db 934 EKERLPQPPCTIDVTYIMVWKCMIDADSRPKELAAEFERNMARDPQRFVITQNRD 993
Qy 1001 GPASPLSTFVRSLEDDMDGLVDAEYLVPOGFCPPDAPGAGGWHHRSSSTRS 1060
Db 994 KLPSPDNSKPFQNLDEEDLEDDMDABEYLV-PAFNIPPP-----IYTSRARIDNSRS 1046
Qy 1061 GGGDLTLGLPSEEEAPRS-----PLAP-SEGAGSDV 1091
Db 1047 -----EIGHSPPPAYTMSGNQFVYRDGGFAAEGVSVYRAPSTTPEAPVQAGATAEI 1101
Qy 1092 FDGLGMAAGKGLQSLTHDPSPLOQYSEDPVPLPS-----ETDGYVAPLTCSPQE 1144
Db 1102 FDDSCCGTLRKVPAPHVQEDSSQYISADPTVAPERSPRGELDESGYTPMRDKPKQE 1161
Qy 1145 YVNOQDVVRPPSPRGPLPAAPAGATLERAKTLPSPGKGVKVDVPAFGAVENPEYLT 1204
Db 1162 YLANPVE-----ENPFVSR-----KNGDLQ-----ALDNPPEYHN 1190
Qy 1205 PQGGAAPQHPHPPA-----FSPAFDNLVYWD 1230
Db 1191 ASNG-----PPKAEDBYVNEPLVNTFANTLGAKEYLKNILSMPEKAKAFDNPYWN 1244
Qy 1231 QDPPERGA--PPSTFKGTPT-----AENPEYL 1255
Db 1245 HSLPPRSTLQHPDLQYLSYTKFYKQNGRIRPIVAENPEYL 1285
RESULT 7
ID_ERB4_RAT STANDARD; PRT: 1308 AA.
AC Q62956; Q622N7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN ERBB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Balliga R.R., Opel D.J., Han X.,
RA Marchionni M.A., Kelly R.A.;
RT "Neuregulins promotes survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes.";
RL J. Biol. Chem. 273:10261-10269 (1998).
RN [2]
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704 (1991).
RN [3]
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659 (1997).
CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS. NRG-
CC 2. NRG-3. HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL). Type I membrane protein.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF041838; AAC03899.1; -.
CC EMBL; U52531; AAC53051.1; -.
CC HSSP; P11362; 1PFGK.
CC InterPro; IPR000494; EGF_R_L domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC InterPro; IPR004019; YLP motif.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF01030; Recep_L_domain; 2.
CC Pfam; PF02757; YLP; 2.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Euk_pkinase; 1.

DR	SMART; SM00261; FU: 4.	169	YODTILWKDIFHKNNQALTLTIDNRSRACHPCSPMKGSRGSCWGESSEDCQSLTRTVCA	QY
DR	SMART; SM00219; TyrKc; 1.	157	YADTIHQDILVRNPWFNMTLSTIGSSGGRCHKCTG-RCWGPTENHCOTLTRTVCAE	Db
DR	PROSITE; PS00107; PROTEIN KINASE ATP; 1.	229	GC-ARCKGFLPTDCCHEQCAAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTPES	QY
DR	PROSITE; PS00109; PROTEIN KINASE TYR; 1.	216	QCDGRCVGPVSDCCHECAGGCGPKDTCFACMFNFDSGACVTCQCPQTFVNTPTFQL	Db
KW	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	288	MNPBGRYFGACVYACVYNYLSTVDGCTLVCPHMQEVTAEADGTQRECKSKPCARV	QY
KW	Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.	276	EHNFNKTYTGAFCKVCKPHNFV-VDSNSCVRACPSKMEV-BENGIMCKPCPTDICKPA	Db
FT	SIGNAL	348	CYGLGMEHLREVRATVSANIOEFAGCKIFGSLAFIPESFDGDPASNTAPLOEQLOVFE	QY
FT	CHAIN	334	CDGIGTGLMSAQTVDSSNIDKFINKTKINGNLIFLVTGHHGDPYNAIDAIDEKLVNR	Db
FT	DOMAIN	408	TLEEITGLYISAWPDSLPLDSVFQNLQVIRGRILHNGAYSLTQGLGISWLGSLRUREL	QY
FT	NP BIND	394	TVREITGFLNIQTWPPNMTDFSVFNLVTIGRVLVSLGLSLILKQOGITSLQFSLKEI	Db
FT	BINDING	468	GSGLALIHNTLHLCFVHTVPWDQLFNPHQALLHTANRDEDECVGSLACHQICARGHCW	QY
FT	ACT SITE	454	SAGNIYITONSNLCTYHTINMTLTFSTVNQRIYIRDNRRAENCTAEGMVCNHLCSNDGCW	Db
FT	DISULFID	528	GPPTQCVNCSQFLRGQECVEECRVLQGLPREVYNARHCLPCHPEQCP-QNGSVTCFGE	QY
FT	DISULFID	514	GPDPQCLSCRRFSRGKICIESCNLYDGEPRFENGISICEVCSQCKQKEDGLTCHGPG	Db
FT	DISULFID	587	ADQCVACAHYKDPFPCVARCPGKPDLSYMPWKPFDEBEGACQCPINCHSCVDLDDK	QY
FT	DISULFID	574	PDNCTCKSHFKDGPNCVEKCPDLQGANSP--IFKVADQDRECHPCHPNCTQCGNGP	Db
FT	DISULFID	647	GC-----PAGORASPLTSIVSAVV-GILLVVVLGVFGILIKRQOKIRKVTMR	QY
FT	DISULFID	632	DCIYYPWTGHSTLPOHAR-TPL--IAAGVIGGFILVIMALTFAVYVRRISIK-KGALR	Db
FT	DISULFID	695	RLAQETELVEPLTPSGAMPNQAORILKETELRKVKVLGSGAGTGVYKGIWIPDGENVKI	QY
FT	DISULFID	688	RFL-ETELVEPLTPSGTAPNQAORILKETELRKVKVLGSGAGTGVYKGIWIPDGENVKI	Db
FT	DISULFID	755	PVAIKVLENTSPKANKETLDEAVMAGVSPVSRLLGICLTSTVQLVTQLMYPYGLLD	QY
FT	DISULFID	747	PVAIKILNETTGPKANVEFMDEALIMASVDHPLHLLVCLSPITQVLTQLMYPHGCLE	Db
FT	DISULFID	815	HYRENRLGSLQDLNWCQIAKMSYLEDVRLVHRDLAARNVLKSPNPKVITDFGLAR	QY
FT	DISULFID	807	YVHEKDNIGSQDLNWCQIAKMSYLEDVRLVHRDLAARNVLKSPNPKVITDFGLAR	Db
FT	DISULFID	875	LLDIDETEHADGKGVPIKMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDCIPA	QY
FT	DISULFID	867	LLEGDEKEYNADGKGVPIKMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDCIPA	Db
FT	DISULFID	935	REIPDLLEKGERLPQPPICITIDVYIMVIMKWMIDSECRPRFRELVSFSSMAEDPQRFV	QY
FT	DISULFID	927	REIPDLLEKGERLPQPPICITIDVYIMVIMKWMIDSECRPRFRELVSFSSMAEDPQRFV	Db
FT	DISULFID	995	IQNED-LGPASPLDSTFYRSLLEDDMDGLVDAAEVLVPOQGFCDPP-----1041	QY
FT	DISULFID	987	IQGDDRMKLPSPNDSKFPQNLLEDDMDAAEVLVLP-QAFNIPPIYTSRTRIDSNR	Db
FT	DISULFID	1042	-----APGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEAGSD	QY
FT	DISULFID	1046	SETHGSPPPAYTWMSSQFVYQGGFATQGG---MMPYTTATSTTPEAPVA--QGATAE	Db
FT	DISULFID	1091	VFDGLDGMGAAGLQSLPHDPSPLQRYSEDPTVPLPS-----ETDGYVAPLTCSPQP	QY
FT	DISULFID	1101	MFDDSCCNGLTRKPVVPHVQEDSSTQRYSDPTVFAPERNPRAELDEEGYMTMHDKPKQ	Db
FT	DISULFID	1144	EYVNPQDVRPQPSPRGGLPLPAARPGATLERAKTLSPGKGVVQDVFAFGGAVENTEYL	QY
FT	DISULFID	1161	EYLNPEV-----ENPFVSR-----KNGDLQ-----ALDNPYH	Db
FT	DISULFID	1204	TPQGAAPQPHPPA-----FSPAPDNLYW	QY

Query Match 42.5%; Score 2911; DB 1; Length 1308;
Best Local Similarity 44.3%; Pred. No. 8.9e-148;
Matches 604; Conservative 190; Mismatches 384; Indels 184; Gaps 30;

1 MELA-ALCRWGLL--ALLPFGASTQVCTDMKRLPASPETHLMDLRLHYGCGQVQ 57
1 MKLATGLVWGSLLVAARTVQFSASVCAGTENKLSLSLEQYRALRYENCEVYM 60

58 GNLELTLYPTNASLSFLQDIEV-----QFNFTVSWLVRPKVSAHSLRQLRV 108
61 GNLEITSNEHRDLSFLRSIEVYGYVLVALNQRYPL-----ENLRIL 105

109 RGTOLFEDNYALAVLNDGDPNLNTPVTGASPGGLRELQRLSLTEILKGGVLIORNPOLC 168
106 RGTKLVEDRYALAIPLNRYKDNF-----GLQELGLKNLTEILGGVYVDQNKPLC 156

F;658-660/Domain: transmembrane #status predicted <TN>
F;723-988/Domain: protein kinase homology <KIN>
F;731-739/Region: protein kinase ATP-binding motif
F;71,191,263,535,576,634/Binding site: carboxylate (Asn) (covalent) #status predicted
F;691/Binding site: phosphate (Thr) (covalent) #status predicted
F;758/Active site: Lys #status predicted
F;882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 86.2%; Score 5860; DB 1; Length 1260;
Best Local Similarity 86.0%; Pred. No. 3.3e-234;
Matches 1085; Conservative 51; Mismatches 109; Indels 16; Gaps 5;

Qy 1 MELAAALCRWGLLLALLPPGAASSTVCTGDMKLRPASPTHLDMLRHLYGCGQVQGNL 60
Db 4 MELAAALCRWGLLLALLPPGIAAGTCTGDMKLRPASPTHLDMLRHLYGCGQVQGNL 63
Qy 61 ELTYLPTNASLSFLQDIEQVGYVLIHNOVQVPLQRLVRGTQLPEDNYALAVLDNG 120
Db 64 ELTYVPAVASLSFLQDIEQVGYVLIHNOVQVPLQRLVRGTQLPEDNYALAVLDNR 123
Qy 121 DPLNNTTPVT-GASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILKXDIHFHKNQL 179
Db 124 DPQDNVAASTPGRTPEGLRELQRLSLTEILKGGVLIQRNPOLCYQDMVLMKDVFRKNQL 183
Qy 180 ALTLIDNRSRACHPCSPMKGSRGWSSSDQSLTRTVGAGCARCKGLPTDCCHEQ 239
Db 184 APVDIDNRSRACHPCAPACKDNHGWSSSDQSLTRTVGAGCARCKGLPTDCCHEQ 243
Qy 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAC 299
Db 244 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTTC 303
Qy 300 PYNLSTDVSGCTLVCPHNOEVAEDGTQRCCKSPCARVCYGLGQVYIKANSKFIGI 359
Db 304 PYNLSTEVSGCTLVCPHNOEVAEDGTQRCCKSPCARVCYGLGQVYIKANSKFIGI 363
Qy 360 TELEFACKKIFGSLAFPLPESFDGDPASNTAPLOPEQLQVETLEETGYLIYISAWPDSL 419
Db 364 NVQBFDFCKKIFGSLAFPLPESFDGDPASNTAPLOPEQLQVETLEETGYLIYISAWPDSL 423
Qy 420 PDLVSFONLQVIRGRIHNGAYSITLQGLGISWGLSLRELGLSLIHHNHLCFVHT 479
Db 424 RDLVSFONLQVIRGRIHNGAYSITLQGLGISWGLSLRELGLSLIHHNHLCFVHT 483
Qy 480 VPMQDLFRNPHQALLHTANRDEDE-CVGEGLACHQLCARGHCWGPPTQCVNCSQFLRGQ 538
Db 484 VPMQDLFRNPHQALLHSGNREEDLCVSSGLVNCNLCAHGHGWGPPTQCVNCSHFLRGQ 543
Qy 539 ECVEECRVLOGLPREYNARHCLCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVA 598
Db 544 ECVEECRVWGLPREYVSDKRLCHPECPQNGSVTCFGEADQCVACAHYKDSSCVA 603
Qy 599 RCPGSGVPLSYMPIWKPDEEGACQPCINCHSCVDLDDKGPASQASPLTSIOYIK 658
Db 604 RCPGSGVPLSYMPIWKPDEEGACQPCINCHSCVDLDDKGPASQASPLTSIOYIK 660
Qy 659 ANSK-----FIGITELKRQKIRKYTMRLLOETELVEPLTPSGAMPNQAWRI 708
Db 661 ATVEGVLFLLLVVVVGIL-IKRRQKIRKYTMRLLOETELVEPLTPSGAMPNQAWRI 719
Qy 709 LKETELRKVKLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVM 768
Db 720 LKETELRKVKLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVM 779
Qy 769 AGVSPYVSRLLGLCLSTVQLVTPMLPYGCLLDHVRNCRGLSQDILLNWCMIAGMS 828
Db 780 AGVSPYVSRLLGLCLSTVQLVTPMLPYGCLLDHVRNCRGLSQDILLNWCMIAGMS 839
Qy 829 YLEVDRLVHRLDAAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNWALESI 888
Db 840 YLEVDRLVHRLDAAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNWALESI 899
Qy 889 LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPDLLEKGERLPQPPICTIDVYMI 948

Db 900 LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPDLLEKGERLPQPPICTIDVYMI 959
Qy 949 MVKCMWIDSECRPRFRELVSERFARMARDPORFVVIQNEIDLGPASPLDSTFVRSLEDDDM 1008
Db 960 MVKCMWIDSECRPRFRELVSERFARMARDPORFVVIQNEIDLGPASPLDSTFVRSLEDDDM 1019
Qy 1009 GDVLDAEYILVPOQGFPCPDPAFAGAGVMHRRSSSTRSGGDLTLGLPSEEEAPRSP 1068
Db 1020 GDVLDAEYILVPOQGFPCPDPAFAGAGVMHRRSSSTRSGGDLTLGLPSEEEAPRSP 1079
Qy 1069 LAPSEGAGSVFDGDLGMRAKGLQSLPHTDPSPLOQRYSEDPTVPLPSETDGYVAPLTCS 1128
Db 1080 LAPSEGAGSVFDGDLGMRAKGLQSLPHTDPSPLOQRYSEDPTVPLPSETDGYVAPLTCS 1139
Qy 1129 PQEYVNPQDVRPQPSREGPLPAARPAAGATLERAKTLSPGKNGVVKDFAFGAVENP 1188
Db 1140 PQEYVNPQDVRPQPSREGPLPAARPAAGATLERAKTLSPGKNGVVKDFAFGAVENP 1199
Qy 1189 EYLTPOGGAAPHPHPPAFSPAFDNLVYWDQDPPERGAPESTFKGTPTAENPEYLGLDVP 1248
Db 1200 EYLVPRGTAAPHPHPPAFSPAFDNLVYWDQDPPERGAPESTFKGTPTAENPEYLGLDVP 1259
Qy 1249 V 1249
Db 1260 V 1260

RESULT 3
I48161
p-185 precursor - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C:Accession: I48161
R:Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa, Gene 140, 251-255, 1994
A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A:Reference number: I48161; MUID:9413007; PMID:7908275
A:Accession: I48161
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1254 <RES>
A:Cross-references: GB:D16295; NID:G493236; PIDN:BAA03801.1; PID:G747595
C:Genetics:
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP
F;718-983/Domain: protein kinase homology <KIN>
F;726-734/Region: protein kinase ATP-binding motif

Query Match 85.8%; Score 5833.5; DB 2; Length 1254;
Best Local Similarity 85.3%; Pred. No. 4e-233;
Matches 1075; Conservative 60; Mismatches 108; Indels 17; Gaps 3;

Qy 1 MELAAALCRWGLLLALLPPGAASSTVCTGDMKLRPASPTHLDMLRHLYGCGQVQGNL 60
Db 1 MELAAALCRWGLLLALLPPGASGTVCTGDMKLRPASPTHLDMLRHLYGCGQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIEQVGYVLIHNOVQVPLQRLVRGTQLPEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIEQVGYVLIHNOVQVPLQRLVRGTQLPEDNYALAVLDNR 120
Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILKXDIHFHKNQL 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILKXDIHFHKNQL 180
Qy 181 LTLIDNRSRACHPCSPMKGSRGWSSSDQSLTRTVGAGCARCKGLPTDCCHEQ 240
Db 181 LTLIDNRSRACHPCSPMKGSRGWSSSDQSLTRTVGAGCARCKGLPTDCCHEQ 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAC 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTTC 300

QY 301 YNYLSTVDGCTLVCLPLHNOEVTAEADGTQRCCKSKPCARVCYGLGMOYIKANSKFIGIT 360
Db 301 YNYLSTVGSTLVCLPLHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLRGARAIYSAN 360
QY 361 ELEFAGCKKIFGSLAFSPESDGPASNTAPLOEQLOVFTLEBEITGYLIYSAMPDPLP 420
Db 361 IQEFAGCKKIFGSLAFSPESDGNPSSGIAPIPTPEQLQVFTLEBEITGYLIYSAMPDSLH 420
QY 421 DLSVFQNLQVTRGRILLUNGASVLTQGLGIGSWGLRSRLRELSGLALIHNTHTLCFVHTV 480
Db 421 DLSVFQNLVTRGRVLDHGDGASLALQGLGIRWGLRSRLRELSGLVLIHRTHTLCFVHTV 480
QY 481 PWDOLFNRPHQALLHTANRPEDECVGEGCLACHOLCARGHCWGPGTQCVCNCSQFLRGQEC 540
Db 481 PWDQLFRNPHQALLHSGNPSEEBECGLKDFACYPICAHGHCWGPGTQCVCNCSHFLRGQEC 540
QY 541 VEECRVLQGLPREYVNAHRLCPHPECPQNGSVTCFQPEADQCACAHYKDDPPFCVARC 600
Db 541 VKECRVMKGLPREYVNGKHCLPCHPECPQNGSTCTGSEADQCTACPHYKDSFFCVARC 600
QY 601 PSGVKPDLUSYMPIWKPDEEGACQPCINCTHSCVDLDDKGCAPAEORASPLTSTQIYKAN 660
Db 601 PSGVKPDLUSYMPIWKPDEEGWCQPCINCTHSCVDLDERGCPAEORASPLTSTSI-----I 655
QY 661 SKFTGI-----TELAKRROKIRKYTMRLLQETELVEPLTPSGAMPNQAQWRIL 709
Db 656 ATVVGILLFLVGVVVGILIKRRKQIRKYTMRLLQETELVEPLTPSGAMPNQAQWRIL 715
QY 710 KETELRKVKVLGSGAGFTVYKGIWPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMA 769
Db 716 KETELRKVKVLGSGAGFTVYKGIWPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMA 775
QY 770 GVGSPYVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRLGSDLLNWCQIAKMSY 829
Db 776 GLGSPYVSRLLGICLTSTVQLVTQMLPYGCLLDHVREHRLGSDLLNWCQIAKMSY 835
QY 830 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETVHADGGKVPYKWWALESIL 889
Db 836 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETVHADGGKVPYKWWALESIL 895
QY 890 RRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTICTIDVYIM 949
Db 896 RRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTICTIDVYIM 955
QY 950 VKCWMIDSECRPRRELVSFSEMRARDPQRFVWQNEIDLGPASPLDSTFYRSLLEDDDMG 1009
Db 956 VKCWMIDSECRPRRELVSFSEMRARDPQRFVWQNEIDLGPSSPLDSTFYRSLLEDDDMG 1015
QY 1010 DLVDAEYLVPOQGFCCPDPAAGGMVHRRHSSTRSRGGDLTLGLEPSEEAAPRSPL 1069
Db 1016 DLVDAEYLVPOQGFCCPDPAAGGTAHRRHSSTRSRGGDLTLGLEPSEEGEPSPRSPL 1075
QY 1070 APSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDTVPLPSTDGYVAPLTCSP 1129
Db 1076 APSEGAGSDVFEGLGMAATGKQPSISPRDLSPQLQRYSEDTPLPTTDCGYVAPLACSP 1135
QY 1130 QPEYVNPQVPRQPPSPREGPLPAARPAGATLERAKTLSPKNGVVKVDVPAFGAVENPE 1189
Db 1136 QPEYVNPQVPRQPPSLTPEGLPLPVRPAGATLERPKTLSPKNGVVKVDVTFGGAVENPE 1195
QY 1190 YLTPQGGGAPOPHPPAPSPAFDNLVYWDQPPPERGAPPSFTKGTPTAENPEYVGLDVPV 1249
Db 1196 YLVPRGGSASQPH-PPALCPAFDNLVYWDQPPSERGSPFNTFEGTPTAENPEYVGLDVPV 1254

RESULT 4

GQHUE

epidermal growth factor receptor precursor - human

Nr:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB

C:Species: Homo sapiens (man)

C:Date: 15-Nov-1984 #sequence revision 27-Nov-1985 #text_change 11-Jun-1999

C:Accession: A00641; A25772; S30024; A38672; A00642; A43615; A23062; A05281; A60143; A33

R;Ullrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.; Y
rg, P.H.
Nature 309, 418-425, 1984
A:Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression of
A:Reference number: A00641; MUID:84219729; PMID:6328312
A:Molecule type: mRNA
A:Residues: 1-1210 <ULL>
A:Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA25240.1; PID:g757924
A:Note: the authors translated the codon AAG for residue 540 as Asn
R;Ishii, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
A:Title: Characterization and sequence of the promoter region of the human epidermal growth
A:Reference number: A25772; MUID:85270438; PMID:2951899
A:Accession: A25772
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-29 <ISH>
A:Cross-references: GB:M11234; NID:g181981; PIDN:AAA52370.1; PID:g553272
R;Haley, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ullrich, A.; Waterfield, M.
Oncogene Res. 1, 375-396, 1987
A:Title: The human EGF receptor gene: structure of the 110 kb locus and identification of
A:Reference number: S30024; MUID:88217333; PMID:3329716
A:Accession: S30024
A:Molecule type: DNA
A:Residues: 1-29 <HA2>
A:Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
R;Haley, J.D.; Waterfield, M.D.
J. Biol. Chem. 266, 1746-1753, 1991
A:Title: Contradictory effects of de Novo transcription and premature transcript terminat
A:Reference number: A38672; MUID:91107677; PMID:1988448
A:Accession: A38672
A:Molecule type: DNA
A:Residues: 1-29 <HAU>
A:Cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271
A:Experimental source: carcinoma cell line A431-7
R;Xu, Y.; Ishii, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.; Merl
Nature 309, 806-810, 1984
A:Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RNAs c
A:Reference number: A00642; MUID:84245835; PMID:6330563
A:Accession: A00642
A:Molecule type: mRNA
A:Residues: 'SCAVRRA', 150-187, 'KSVIQAV', 195, 'M', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-321
, 798-799, 'TD', 802-811, 'R', 813-942 <XUY>
A:Experimental source: A431 human carcinoma cells, which have large numbers of EGF receptor
R;Lin, C.R.; Chen, W.S.; Krueger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verma, I.M.
Science 224, 843-848, 1984
A:Title: Expression cloning of human EGF receptor complementary DNA: gene amplification;
A:Reference number: A43615; MUID:84196372; PMID:6326261
A:Accession: A43615
A:Molecule type: mRNA
A:Residues: 713-964 <LIN>
A:Experimental source: epidermoid carcinoma cell line A431
R;Simmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.
Biochem. Biophys. Res. Commun. 124, 125-132, 1984
A:Reference number: A23062; MUID:85046483; PMID:6093780
A:Accession: A23062
A:Molecule type: mRNA
A:Residues: 1028-1210 <SIM>
R;Weber, W.; Gill, G.N.; Speiss, J.
Science 224, 294-297, 1984
A:Reference number: A05281; MUID:84172183; PMID:6324343
A:Accession: A05281
A:Molecule type: protein
A:Residues: 25-30, 'S', 32-51, 454-467 <WEB>
R;Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.
J. Biol. Chem. 260, 5205-5208, 1985
A:Title: Identification of residues in the nucleotide binding site of the epidermal growth
A:Reference number: A60143; MUID:85182650; PMID:2985580
A:Accession: A60143
A:Molecule type: protein
A:Residues: 740-744, 'X', 746-747 <RUS>
R;Mroczkowski, B.; Mosig, G.; Cohen, S.

Nature 309, 270-273, 1984
 A>Title: ATP-stimulated interaction between epidermal growth factor receptor and superoxide dismutase
 A;Reference number: A38023; MUID:84191554; PMID:6325948
 A;Contents: annotation; receptor activity
 R;Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.; Cell 59, 33-43, 1989
 A>Title: Functional independence of the epidermal growth factor receptor from a domain in the extracellular region
 A;Reference number: A33331; MUID:90003233; PMID:2790960
 A;Contents: annotation; internalization signal
 C;Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor complex
 C;Genetics:
 A;Gene: GDB:EGFR
 A;Cross-references: GDB:120610; OMIM:131550
 A;Map position: 7p12.3-7p12.1
 C;Superfamily: epidermal growth factor receptor; protein kinase homology
 C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphotyrosine
 F;1-24/Domain: signal sequence #status predicted <SIG>
 F;25-1210/Product: EGF receptor #status predicted <NAT>
 F;23-645/Domain: extracellular #status predicted <EXT>
 F;75-300/Domain: EGF receptor extracellular domain repeat <EE1>
 F;390-600/Domain: EGF receptor extracellular domain repeat <EE2>
 F;646-668/Domain: transmembrane #status predicted <TM>
 F;669-1210/Domain: intracellular #status predicted <INT>
 F;710-975/Domain: protein kinase homology <KIN>
 F;718-726/Region: protein kinase ATP-binding motif
 F;999-1046/Region: coated-pit mediated internalization signal
 F;1047-1210/Region: inhibitory
 F;128,175,352,413,444,528,603/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;745/Active site: Lys #status experimental

Query Match 45.8%; Score 3115; DB 1; Length 1210;
 Best Local Similarity 49.5%; Pred. No. 3.2e-121;
 Matches 624; Conservative 174; Mismatches 359; Indels 104; Gaps 21;
 QY 11 LLLALLPGAA--STVCTGTDMLRLPASPTHLDMLRLHYQQCVQVGNLEITYPTN 68
 DB 14 LLAALCPASRALEKKVCCQGSNKLTLQGTEDFLSLQRMFNCEVVLGNLEITYQVN 73
 QY 69 ASLSFLDIOIQVQGVLIHANOVQVPLQRLURIVRGTLQEDNYALAVLDNGDPLNNTTP 128
 DB 74 YDLSFLKTIQVAGVLIALTVERIPLENLIQIRGNMYENSVALVLSNYD----- 126
 QY 129 VTGASPGGLRELQRLSLEILKGGVLIQORNQLCYQDTILMKDIFHRKNQALFLIDTNR 188
 DB 127 ---ANKTGLKELPMRNLEQLHGAVRFESNNPALCNVESIQWRDIVSSDFLSNMSMDFNH 183
 QY 189 SRACHPCSPMKSGRCSWGESSEDCQSLTRTVACGGCA-RCKGPLPTDCCHQCAAGCTGP 247
 DB 184 LGSQCKDPSCPNGSCWAGAGENCQKLTIIQAQCSGRCKRGSFSDCHNQCAAGCTGP 243
 QY 248 KHSCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPVNVLST 307
 DB 244 RESCLVCRKFRDPAKCTCPPLMLNPTTYQMDVNPGEKYSFGATCVKCKPRNYVYTD 303
 QY 308 VGSCTLVCLPHNQVTAEDGTQRCEKSKPCARVCYGLGMQYIKANSKFIGITILE-FAG 366
 DB 304 HGSCVRACGADSYEM-BEDGVKCKCEGPKRVCNGIGIGEFK-DSLSINATNWKFKN 361
 QY 367 CKKIFGSLAFIPESFDGDPASNTAPLOEQLOVFETLEETGYLYISAWPDSLDLSVQ 426
 DB 362 CTSISGDLHLPVAFRGDSFTHTPPLDPQELDIKTVKEITGFLLIQAWPENRTDLHAFE 421
 QY 427 NLOVIRGRILHNGAYSLLTQGLGSLWGLSLRLSGLSLALIHHTHLCFVHTVPWDLF 486
 DB 422 NLEIRTKHQGFSLAVSLNLTISGLRSLKISDGDVLIISGNKLVCANTINWKFLF 481
 QY 487 RNPQALLHTANRPEDCEVCEGLACHQICARGHCWGPFGPTQVNCQSQFLRGQECVERCV 546
 DB 482 GTSSQKTKIISNRGENSKATGVCHALCSPEGCWGPEDPCVSCRNVSRGRCVCKKL 541
 QY 547 LQGLPREVYNARHCLPCHPECPQNGSVTCFPGPADQCVACAHYKDPFCVACPSGVKP 606

DB 542 LEGEPREFVENSECICHPECLFOAMNITCTGRPDNCIOCAHYIDGPHCVKTCPCAGVMG 601
 QY 607 DLSYMPIWKFDEDEGACQCPINCTHSCVDLDKGCPCAEORASPLTSIQVIKANSKF--- 663
 DB 602 ENNTL-VWYADAGHVCHLCHPNCYCTCTGPGEGCTNGPKTSPSIATGMVALLLVV 660
 QY 664 -IGITELRRQOKIRKYVMRLLOETVELPLTPSGAMPNOAQMRLKETELRKVKVLGS 722
 DB 661 ALGIGLPMRRRHVVRKTRLLQERLEVLPLTPSGEAPNQAALLRIKETEFKKIKVLGS 720
 QY 723 GAGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVSGPYVSRLLGI 782
 DB 721 GAGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVSGPYVSRLLGI 780
 QY 783 CLTSTVOLTLMPYGLLDHVRNRLGSGQDILLNMCQIAKMSYLEDVRLVHRDLAA 842
 DB 781 CLTSTVOLTLMPYGLLDHVRNRLGSGQDILLNMCQIAKMSYLEDVRLVHRDLAA 840
 QY 843 RNVLVKSPNHVKITDFGLARLLDIDETEHADGGKVPDKWMALESILRRRFTTHQSDVMSY 902
 DB 841 RNVLVKTPQHKITDFGLARLLDIDETEHADGGKVPDKWMALESILRRRFTTHQSDVMSY 900
 QY 903 GVTVMELMTCAKPYDGIIPAREIPDLLEKGERLPDPDICTIDYVMVIMVKWMDSECRP 962
 DB 901 GVTVMELMTCAKPYDGIIPAREIPDLLEKGERLPDPDICTIDYVMVIMVKWMDSECRP 960
 QY 963 FRELVSFNRNARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDMDGLVDABEYLVPQ 1021
 DB 961 FRELVSFNRNARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDMDGLVDABEYLVPQ 1020
 QY 1022 QGFFCPDPAFCAGGMVHRRSSTRSGGGDLTLGLEPSEEAAPRSPAPSEAGSDVFD 1081
 DB 1021 QGFFCPDPAFCAGGMVHRRSSTRSGGGDLTLGLEPSEEAAPRSPAPSEAGSDVFD 1080
 QY 1082 GDLMGNAKGLQSLPHDPSLQRYSEDPTVPLPSET--DGYVAPLTCSPQEVNQPVDV 1139
 DB 1045 STVACIDRNLGSCPTKEDSFQRYSDPTGALTEDSIDDTFL-----VPVEYINQ-SV 1097
 QY 1140 RPOPPSPREGLPAAPAPAGATLERAKTSLPKGVKVDVFAFGAVENPVL-TPQGGAA 1198
 DB 1098 PKPAGSVQNVVHNPQNP-----APSRDPHQD--PHSTAVGNPEYLVNQ---- 1143
 QY 1199 PQHPPPAFSPAFDNLVYWDQ-----DP-----PERGAPPTFKGTPTAENPEY 1242
 DB 1144 ----PTCVNSTFSDPAHWAQKSHOISLDNPDYQDDFFPKKAKPNCIFKGS-TAENAEY 1197
 QY 1243 L 1243
 DB 1198 L 1198
 RESULT 5
 A53183
 epidermal growth factor receptor precursor - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999
 C;Accession: A53183; A43818; S24942; A28941; S45325; I49543
 R;Luetteke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N.A.; Genes Dev. 8, 399-413, 1994
 A;Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor
 A;Reference number: A53183; MUID:94170986; PMID:8125255
 A;Accession: A53183
 A;Molecule type: mRNA
 A;Residues: 1-1210 <LUE>
 R;Avivi, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B. Oncogene 6, 673-676, 1991
 A;Title: Comparison of EGF receptor sequences as a guide to study the ligand binding site
 A;Reference number: A43818; MUID:91232866; PMID:2030916
 A;Accession: A43818
 A;Molecule type: mRNA
 A;Residues: 1-714 <AVI>
 A;Cross-references: GB:X59698

R;Eisinger, D.P.; Serrero, G.
 submitted to the EMBL Data Library, June 1992
 A;Reference number: S24942
 A;Accession: S24942
 A;Molecule type: mRNA
 A;Residues: 969-971,'K',973-1115,'D' <EIS>
 A;Cross-references: EMBL:Z12608
 R;Heiseremann, G.J.; Gill, G.N.
 J. Biol. Chem. 263, 13152-13158, 1988
 A;Title: Epidermal growth factor receptor threonine and serine residues phosphorylated in
 A;Reference number: A28941; MUID:88330814; PMID:3138233
 A;Accession: A28941
 A;Molecule type: protein
 A;Residues: 689-694,'X',696-704,'L',706-707;989-992,'XX',995-996,'X',998-1000;1002-1009,
 R;Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
 submitted to the EMBL Data Library, April 1994
 A;Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
 A;Reference number: S45325
 A;Accession: S45325
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-971,'K',973-1210 <VER>
 A;Cross-references: EMBL:X78987; NID:9488830; PIDN:CAA55587.1; PID:9488831
 R;Paria, B.C.; Das, S.K.; Andrews, G.K.; Dey, S.K.
 Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
 A;Title: Expression of the epidermal growth factor receptor gene is regulated in mouse b
 A;Reference number: I49643; MUID:93126380; PMID:7678348
 A;Accession: I49643
 A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 12-20,22-132 <RES>
 A;Cross-references: GB:L06864; NID:g193001; PIDN:AAA53029.1; PID:g567201
 C;Genetics:
 A;Gene: EGFR
 C;Superfamily: epidermal growth factor receptor; protein kinase homology
 C;Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop
 F;1-24/Domain: signal sequence #status predicted <SIG>
 F;648-670/Domain: transmembrane #status predicted <TM>
 F;712-977/Domain: protein kinase homology <KIN>
 F;720-728/Region: protein kinase ATP-binding motif
 F;680,695/Binding site: phosphate (Thr) (covalent) #status experimental
 F;697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
 F;993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
 F;1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
 F;1157/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 45.4%; Score 3082; DB 2; Length 1210;
 Best Local Similarity 49.5%; Pred. No. 7.3e-120;
 Matches 630; Conservative 164; Mismatches 362; Indels 116; Gaps 25;

Qy	11	LLLLLPPGAA--STQVCTGTDMLRLPASPEETHLDMLRLHYQGVQVQGNLELTPLTN	68
Db	14	LLTALCAAGGALEKKVCGQTSNRLTQLGTFEDHFLSLQRMYNNECVVLGNLEITYQVRN	73
Qy	69	ASLGFLODIQEVQGVLLAHNOVRQVPLQRLIRVGTOLFEDNVALAVLDNGDPLNNTTP	128
Db	74	YDLFLKTIQEVAGVLLALNTVERIPLENQITGNALYENTYALALSN-----	124
Qy	129	VTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDI----FHRNQALFLI	184
Db	125	-YGNRTGLRELPMNLQELLIGAVRFNSNPFLCNMDTIQWRDINVQVFMNMSMDL---	180
Qy	185	DTNRSRACHPCSPMKGRSGRWESSEDCQSLTRTVACGGCA-RCKGPLPTCCHEQCAAG	243
Db	181	-QSHPSSCPCKDPSPNGSCWGGGEENCOKLTKIICAQCCSHRCGRSPSCCHNQCAAG	239
Qy	244	CTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPNY	303
Db	240	CTGPRESDCLVCQKQFDEATKDTCPPLMLNPTYQMDNPGEKYSFGATCVKCKPNY	299
Qy	304	LSTDVGSCTLVCPPLHNQBVTAEDGTQRCCKSKPCARCYGLGMQYIKANSKFIGITELE	363
Db	300	VVTDHGSVCACGPDYYEV-BEDGIRKCKKCDGPRCKVCNGIGIGEFK-DTSLINATNIK	357

364 -FAGCKKIFGSLAFLPESFDGDPASNTAPLQBPOLQVFEETLEBITGYLIYSAMPDLSPLD 422
 358 HFKYCTAISGDHLHLPVAFKGDSTRTPLDPRELEILKTVKEITGELLQAMPDNWTDL 417
 423 SVFQNLQVIRGRILHNGASLTLLQGLISWLGSLRSLRELGSGLALIHNNHLCFVHTVPW 482
 418 HAFENLEIIRGRTHQHGQFSLAVVGLNITSGLRSLKEISDGDVLIISGNRLCVANTINW 477
 483 DQLFNPHOALLHTANPEDECVEGEGLAHQLCARGHCWGPGTQCVNCQSLRQSCVE 542
 478 KKLFGTNPQKTKIMNNAEKDKAVNVHVCNPLCSSEGCWGPEDRDCVSCQNVSRGECVE 537
 543 ECVLQGLPREYNARHCLCHPECOFQNGSVTCFGEADQCVACAHYKDPFFCVARCP 602
 538 KCVILEGEPREFVENSEICIOCHPECLQAMNICTGRPCNQCQAHYIDGPHCVKTCPA 597
 603 GVKPDLSYMPKIPFPDEBEGACQPCPCINCTHSCVDLDDKCPAEQORASPLTSIQY- 661
 598 GINGENNTL-VMKYADANNVCHLCHANCTYGCAGPGLQGCVEWPSGPKIPSIATGIVGGL 656
 662 KFI-----GITELKRRQOKIRKVTMRLLQETELVEPLTSPGAMPNQAOHRIKTELK 716
 657 LFTVVALGIGLFMRERRHIVKRTLRLLQERELVEPLTSPGAMPNQAOHRIKTELK 716
 717 VKVLGSGAGFTVYKGIWIPGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGSYV 776
 717 IKVLGSGAGFTVYKGIWIPGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGSYV 776
 777 SRLIGICLTSTVQLTQMPYGLLDHVRNRRGLSGQDLNLMCMQIAKMSYLEDVRLV 836
 777 CRLLIGICLTSTVQLTQMPYGLLDHVRNRRGLSGQDLNLMCMQIAKMSYLEDVRLV 836
 837 HRDLAARNVLKSPNVKVIITDFGLARLLDIDETEHADGKVPKIKWMALESILRRRTHQ 896
 837 HRDLAARNVLKSPNVKVIITDFGLARLLDIDETEHADGKVPKIKWMALESILRRRTHQ 896
 897 SDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVYMIWVKW 956
 897 SDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVYMIWVKW 956
 957 SECRPREFRLVSEFNRMDPQRFVVIQ-NEDLGASPLDSTFYRSLLEDMDGDLVDAE 1015
 957 ADSRKPREFLILFESQWARPQVLYIQGDERMHLPSPTDSNFYRALMDEEDMEDVVDAD 1016
 1016 EYLVPOQGFCDPAPGAGGVHRRSSSTRSGGDLTLGLEPSEEAERPSPLAPSEGA 1075
 1017 EYLIPQOGFF-----NSPST-----SRTPLLSLSA 1042
 1076 GSDVFDGDLGMAAKGLQSLPTHDPSPQRYSEDPVLPSET--DGVPAPLTCSPOPEY 1133
 1043 TSN-----NSTVACINRNGSCRVKEDAFQRYSDPTGAVTEDNIDDAFL-----PVPEY 1092
 1134 VNQPDVPRQPPSPREGPLPAARPAAGATLRAKTLSPGKGVVQKDVAFGGAIVENPEYL-T 1192
 1093 VNQ-SVPEKRPAGSVQNFVHNQPLHP-----APGRDLHYQN--PHSNAVGNPEYLT 1141
 1193 PQGAAPQPHPPAFSPAFDNLYWDO-----DP-----PERGAPPSTFKGTPT 1236
 1142 AQ-----PTCLSSGFSNPAWIKGSHQMSLDNPDYQODFPFKETKPNGIFKG-PT 1191
 1237 AENPEYLGLDVP 1248
 1192 AENAEYLRVAPP 1203

RESULT 6
 TVCHLV
 epidermal growth factor receptor precursor - chicken
 N;Contains: protein-tyrosine kinase (EC 2.7.1.12) erbB
 C;Species: Gallus gallus (chicken)
 C;Date: 28-Feb-1996 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
 C;Accession: A27720; A00643

C;Keywords: ATP; growth factor receptor
F;716-981/Domain: protein kinase homology <KIN>
F;724-732/Region: protein kinase ATP-binding motif

Query Match 43.0%; Score 2925.5; DB 2; Length 1308;
Best Local Similarity 45.2%; Pred. No. 2.2e-113;
Matches 608; Conservative 175; Mismatches 385; Indels 177; Gaps 28;

```
Qy 9 WGLLLALLPPGAA-----STQCTGTDMLRLPASPEHLDMLRLHYOCQVVOGNLEITY 64
Db 8 WWSVLLVAAGTVQPSDSQSCAGTENKLSLSLEQQYRALRYKYENCEVVMGNLEITS 67
Qy 65 LPTNASLSFDIOIQVOGVYLIAHNQVQPLRLRIRVGTQPLFEDNYALAVLNGDPLN 124
Db 68 IEHNRDLSFSLRVRETVGVYVALNQFRLPLENLRIIRGTYKLYEDRVALAIFLNYKDG 127
Qy 125 NTPVTGASPGGLRELQLRSITELKGVLIQRPOLCYQDTILWKDIFHKNNQALATLI 184
Db 128 NF-----GLQELGLKNLTELINGGVYVDQNKFLCYADTIHQDIVRNPWPSNLTIV 178
Qy 185 DTRSRACHPCSPMKGSRGSESSDCOSTRTVTCAGGC-ARCKGPLPTDCHEQCAG 243
Db 179 STNGSSGGRCHKSTG-RCWGPTENHCOTLRTVCAEQDGRCYGYPVSDCCHECAGG 237
Qy 244 CTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPNY 303
Db 238 CSGPKDTCFACMFNDSGACVTCQPTFVYNPTTFQLEHFNNAKYTFAGFCVKKCPNF 297
Qy 304 LSTDVGSCTLVCPHLNQHVTAEDTQRCCEKSKPCARVCYGLGMQYIKANSKFIGITE 363
Db 298 V-VDSSSCVACPSKMEV-EENGIMCKPCTDICKPKACDGTGSLMSAQTVSSNIDK 355
Qy 364 FAGCKIIGSLAFLPESFDGDPASNTAPLOEQLOVPTLEITCYLYISWPSLDPLS 423
Db 356 FINCTKINGNLIPLVTGTHGDPYNAIEADPEKLVNFTVREITGFLNQISWPNMTDFS 415
Qy 424 VFQNLQVIRGRILHNGAYSLTLQIGISWGLRSLRELGSGLALHHTHLCFVHTVPWD 483
Db 416 VFSNLVTIGRVLVYSGLSLLILKQOITSLOFOSLKEISAGNIYITDNSLCYHTINWT 475
Qy 484 QLFRNPHQALLHTANRPEDECVGEGLACHQICARGHCWGPQTCVNCQSQFLRGOECVE 543
Db 476 TLFSTINQIRIVTRDNRKAENCTAEGMVCNHLCSGDCGWPQDQCLSCRRESRGICIES 535
Qy 544 CRVLQGLPREYVNAHCLPCHPEQOP-QNGSVTCGPEADOCVAHYKDPFCVARCP 602
Db 536 CNLYDGEPRFENGISICEVCDPQCEKMDGLTCHGPGPDNCTKSHFKDGPNCVEKCPD 595
Qy 603 GVKPDLSTYMPIWKFPDEGACQPCPINCTHSCVOLDDKGC-----PAEORASPL 651
Db 596 GLOGANSF--TFKYADPDECHPCNPCTQCGNGFTSHDCIYYPWTGHTILPOHAR-TPL 652
Qy 652 TSIQVIKANSKFI-----GIT--ELKRROOKIRKYTMRELLQETELVEPLTPSGAMPNOAQ 705
Db 653 IAAAGVI--GGLFILVIGLTFEAVYVRRKSIKKRALRRFL-ETELVEPLTPSGTAPNOAQ 709
Qy 706 MRILKETELRKVKVLGSAFGTVYKGIWIPGENVKIPIVAIKVLRENTSPKANKEILDEA 765
Db 710 LRILKETELRKVKVLGSAFGTVYKGIWIPGETVKIPIVAIKILNETTGPKANVEFND 769
Qy 766 YVMAGVGSPPYVSRILGICLTSTVOLVQLMPYGCILLDHVRNRRGLGSQDLLNWCMTAK 825
Db 770 LIMASMDHPHLVRLLVGLVCLSTIQLVTQLMPHGCLLEYVHEKONIGSQLLNWCVQIAK 829
Qy 826 GMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHYHAGGKVPKKNMAL 885
Db 830 GMYLEERLVRDLAARNVLKSPNHVKITDFGLARLLDEGEKYEYNADGKMPKKNMAL 889
Qy 886 ESILRRRTHOSDVWSYGVYTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDV 945
Db 890 ECIHVRKTHOSDVWSYGVYTWELMTFGKPYDGIPTREIPDLLEKGERLPQPICTIDV 949
Qy 946 YIMVVKCWMIDSECRPRELVSEFSRMAPDQRFVVTQNE-DLGPASPLDSTFVSILLE 1004
```

```
Db 950 YIMVVKCWMIDADSRPKFKELAAEFSSRMARDPQRYLVIQGDDRMKLPSPNDSKFFQNLDD 1009
Qy 1005 DDMGDLVDAAEYLVPOOGFFCDDPAPGAGVMVHRRSSSTRSGGDLTLGLPSPSEEA 1064
Db 1010 EEDLEDMDAAEYLVLP-QAFNIPPP-----IYTSRARIDSNRS-----EIGHSPPPAYT 1057
Qy 1065 PRS-----PLAP-SEGAGSDVDFDGLDGLGGAAGKQLSL 1095
Db 1058 PMSGNQFYVRDGGFAABOQGVSVYRAPSTSIPEAPVAAQATAEIFDDSCCNGTLRKVPAP 1117
Qy 1096 PTHDPSFLORYSESDPTVPLFS-----ETDGVYVAPLTCSPQPEYVNPQDVRPPOPPSPRE 1148
Db 1118 HVQEDSSTQYSADPTVFAPERSPRGELDEEGYTMPWRDKPKQEYLNPE------E 1168
Qy 1149 GPUPAARPAAGATLERAKTILSPGKNGVVKDVPFAGGAIVENPEYLTPOGGAAPQHPPPA-- 1206
Db 1169 NPFVSR-----KNGDLQ-----ALDNPYHNASNG-----PPKAED 1200
Qy 1207 -----FSPAFDNLVYWDODDPPERGA--PPSTFK 1232
Db 1201 EYVNEPLYLNTFANTLGAKEYLKNILMSPEKAKAFDNPDYWNHSLPPRSTLQHPDYLO 1260
Qy 1233 GTPT-----AENPEYL 1243
Db 1261 EYSTKYFYKQNGRIRPIVAENPEYL 1285

RESULT 8
S06142
protein-tyrosine kinase (EC 2.7.1.112) mrk-y precursor - southern platyfish
N;Alternate names: epidermal growth factor receptor homolog; kinase-related transforming
C;Species: Xiphophorus maculatus (southern platyfish)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
C;Accession: S06142; S13809
R;Wittbrodt, J.; Adam, D.; Malitschek, B.; Maeueller, W.; Raulf, F.; Telling, A.; Roberts
Nature 341, 415-421, 1999
A;Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu loc
A;Reference number: S06142; MUID:90015140; PMID:2797166
A;Accession: S06142
A;Molecule type: DNA
A;Residues: 1-1166 <MIT>
A;Cross-references: EMBL:X16891; NID:G65290; PIDN:CAA34770.1; PID:G65291
R;Adam, D.; Maeueller, W.; Scharlt, M.
Oncogene 6, 73-80, 1991
A;Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xiphophoru
A;Reference number: S13807; MUID:91125882; PMID:1846957
A;Accession: S13809
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>
A;Cross-references: EMBL:X56319; NID:G65284; PIDN:CAA39763.1; PID:G65285
C;Genetics:
C;Gene: mrk
A;Map position: Y
A;Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein; tyro
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>
F;707-972/Domain: protein kinase homology <KIN>
F;715-723/Region: protein kinase ATP-binding motif

Query Match 38.8%; Score 2635; DB 1; Length 1166;
Best Local Similarity 45.0%; Pred. No. 1.8e-101;
Matches 570; Conservative 160; Mismatches 393; Indels 144; Gaps 30;

Qy 4 AALRCWGLLALLPPGAAS----OVCTGTDMLRLPASPEHLDMLRLHYOCQVVOGN 59
Db 8 AALLQ--LLLVLSISRCCSTDPDRKVCQGSNTMTM---LDNHYLKKKKYSGCNVLEN 62
Qy 60 LELTYLPTNASLFLQDIOEVOGVYLIHAHQVQRPQLRLIRVGTQLFEDNYALAVLDN 119
```


Db 63 LEITYTQENQDLSEFQSTQEVGGVLIAMNEVSTIPLVNLIRLQNLQYEGNFILLVMSN 122
Qy 120 GDPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQL 179
Db 123 YQK-NPSSP--DVYQVGLKQLQLSNLTELSSGVKSHNPLLCNVEILNWDIVDKTSNP 179
Qy 180 ALTLIDNRSRACHPCSPMKSGRCWGSSESDCSLRTVTCAGGC-ARCKGPLPTDCCHE 238
Db 180 TMNLI PHAFERQCKQCHGCVGSCWAPGPGHCQKFTKLLCAEQCNRRRCRGPKEIDCNE 239
Qy 239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTA 298
Db 240 HCAGCTGPRADCLACKDFNDGTCCKTCTPPKIYDIVSHQVNDPNIKTFGAACVKE 299
Qy 299 CPYNYSLDVGSCTLVCPHLHQVEYTAEDGTQRCCKSPCARVCYVGLGMQVIKANSKF 358
Db 300 CPSNVVTE-CACVRSKSGAGMLEVD-ENGKSRKPCDGVCPKVCDDIGIGSL-SNTIAVN 356
Qy 359 ITEL-EFAGCKKIFGSLAFLEPSTGDPASNTAPLOPQLQVFTLEBITGYLYISAWPD 417
Db 357 STNIRSFNCTKINGDIILNRNSPEGDPHYKIGTMDPEHLNWLITVKEITGYLYVIMWPE 416
Qy 418 SLPDLVFPQNLQVIRGRILHNGAYS-LTLOGLGISWLGRLSLRELGSGLALIHENTHLCF 476
Db 417 NMTSLVFPQNLQVIRGRILHNGAYS-LTLOGLGISWLGRLSLRELGSGLALIHENTHLCF 476
Qy 477 VHTVPWDLFRPHQALLHTANRDECEVGBGLACHQLCARGHGWGPGTQCVNCSQFLR 536
Db 477 ANTINWRFLRSEDSIEYDART-----ENQTCNNECSESDGCM-PGPTMCVSLHVDR 528
Qy 537 GQCVBECEVRLQGLPREVYNARHCLPCHPECPQNGSVTCFPGBADQVCAHYKDDPPFC 596
Db 529 GGRCVASCNLLQGBPREAQVQDRCVQCHOECLVQDLSLTCYGPNGPANGSKSAHFQDGPQC 588
Qy 597 VARCPGKVPDLSYMTKWPDEEGACOPCPINCHTSCVDLDDKCPAE-----QRASPL 651
Db 589 IPRCPHGLGDDTL-INKYADKMGQCPQCHQNTQCGSGGLSGCRGDIVSHSSLAGVL 647
Qy 652 TSIQVIKANSFIFGITELKRQKIRKYTMRLRLQETELVEPLTPSGAMPNQAOIRLKE 711
Db 648 VSGLLITVIVALLIVLLRRRIK-RKETIRCLQEKELVEPLTPSGAQNPAFLRLKE 706
Qy 712 TELRKVKVLSGAGTGVYKGIWIPDGENVKIPVAKVLRNTPSKANKEILDEAYVWAGV 771
Db 707 TEFKDKRVLGSGAGTGVYKGLWNPDPENIRIPVAKVLRNTPSKANKEILDEAYVWASV 766
Qy 772 GSPVSVRLGLTSTVOLVTQMPYGCCLLDHVRNCRGLSGODLWNCWQIAKGMVLE 831
Db 767 DHPVHVRLLGLTSAVOLVTQMPYGCCLLDYVRQHOERICQWLLNWCVQIAKGMVLE 826
Qy 832 DVRLVHRDLAARNVLKSPNPKITDFGLARLLDIDETEHADGKVPKKNMALESILRR 891
Db 827 ERHLVHRDLAARNVLLKNPNHVKITDFGLSKLLTADKEKYQADGKVPKKNMALESILOW 886
Qy 892 RPTHQSDVMSYGVTVWELMTGAKPYDIPAREIPDLLEKGERLPQPPICHTIDYIMIVK 951
Db 887 TYTHQSDVMSYGVTVWELMTGSKPYDIPAKEIASVLENGERLPQPPICHTIEYVMIK 946
Qy 952 CWMIDSECRPRELVEFSRWARDPQFVVIQNEIDLGPASPLDSTFVRSILLEDDMDGL 1011
Db 947 CWMIDPSRRPRELVEFSQWARDPSRYLIQG---NLPSLSDRRFLSRLLSDD--DV 1001
Qy 1012 VDABEYLVPQGGFFCPDPAPGAGGVVHRRSSSTRSGGDLTLGLPSEEAAPRSLAP 1071
Db 1002 VDAEYLLPYKRI-----NRQS-----ECIP 1024
Qy 1072 SEGAGSVDFDGLGMAAKGLQSLPTHDPSPLOYSEDPV-PLPSETDGYVAPLTCSPQ 1130
Db 1025 PTGH-----PVRENSITLRNISDPTQNALEKOLDGH-----1055
Qy 1131 PEYNQDPVRPOP-----PSPRE-----GPLP-AARPAGATLERAKTSLFGKNGVVKD 1177
Db 1056 -EYVNPQGETSRLSDIYNPNYEDLTDGWPVSLSSQEAETNFSRPEYLTNQNSL---1111

Qy 1178 VFAFGGAVENPEYLTPQGAAPQHPHPAPAFDNLNLYWDQPPPERGAPPESTFKGTPTA 1237
Db 1112 PLVSSGSMDDPDY--QAG-----YQAAF-----LPQTGALTNGMFLPAA 1149
Qy 1238 ENPEYLG 1244
Db 1150 ENLEYLG 1156
RESULT 9
A36223
kinase-related transforming protein (erbb3) (EC 2.7.1.-) precursor - human
C:Species: Homo sapiens (man)
C:Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C:Accession: A36223; I59164
R:Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A:Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal g
A:Reference number: A36223; MUID:90083234; PMID:2687875
A:Accession: A36223
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1342 <KRA>
A:Cross-references: GB:M29366
R:Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, G.J
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A:Title: Molecular cloning and expression of another epidermal growth factor receptor-
A:Reference number: I59164; MUID:90311312; PMID:2164210
A:Accession: I59164
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-559, 'G', '561-957', 'F', '959-1063', 'G', '1065-1342 <RES>
A:Cross-references: GB:M34309; NID:gi183990; PIDN:AAA35979.1; PID:g306841
C:Genetics:
A:Gene: GDB:ERBB3; HER3
A:Cross-references: GDB:I19880; OMIM:190151
A:Map position: 12q13-12q13
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
C:Keywords: Atp; phosphotransferase
F:707-972/Domain: protein kinase homology <KIN>
F:715-723/Region: protein kinase Atp-binding motif
Query Match 35.3%; Score 2398.5; DB 2; Length 1342;
Best Local Similarity 40.4%; Pred. No. 1.1e-91;
Matches 530; Conservative 189; Mismatches 456; Indels 137; Gaps 33;
Qy 10 GLLALLPPGAA--STQVCTGTDKMLRLPASPETHLMDLRHLVYGGCOVVOGNLELTYLPT 67
Db 11 GLLFSLARGSEIVGNSQAVCPGTLNGLSVTDAENQVQTLKLYERCEVVMGNLEIVLTGH 70
Qy 68 NASISFLQDIQEVQGVVLIHNOVRQVPLQRLRIVRGTLQFEDNYALAVLDNGDPLNNTT 127
Db 71 NADLSFLQWIREVTVGVVYLVAMNEFSTLPLNLRVVRGTQVYDGKFAIFVM-----LNYNT 125
Qy 128 PVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQALTLIDTN 187
Db 126 ----NSSHALRQLRLTOLTEILSSGVYIEKNDKLCHMDTIDWRDIVDRD---AEIVVKD 178
Qy 188 RSRAHPSPMKSGRCWGSSESDCSLRTVTCAGGC-ARCKGPLPTDCHEQCAAGCTG 246
Db 179 NGRSCPCHVCKG-RCWGPGSEDCOTLTKTICAQCNHGHCFGNPNQCCHECAGCGSG 237
Qy 247 PKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTCAPYNYLST 306
Db 238 PQDTDCPACRHFNDSGACVPRCPQPLVYNKLTQLEPNPHTKYQYGGVCVASCNPNFV-V 296
Qy 307 DVGSCTIVCLPHNQVEYTAEDGTQRCCKSPCARVCYVGLGMQVIKANSKF--IGITELE- 363
Db 297 DQTSVPRACPDKMEVD-KNGLKWCPCGGLCPKACEGTG-----SGSRQTVDSSNIDG 350
Qy 364 FAGCKKIFGSLAFLEPSTGDPASNTAPLOPQLQVFTLEBITGYLYISAWPDSPLDS 423

351	Db	FVNCTKILGNLDFLITGLNGDPWHKIIPALDPEKUNVFRVTREITGYLNIQSWPPHMNF	410
424	Qy	VFQNLQVIRGRILHNAYS-LITLOGLGTSWLGRLSRELASGLALIHHTHLCFVHTVPV	482
411	Db	VFSNLTTIGRSLYNRGFSLLIMKNLVNTSLGFSLSKEISAGRIYISANRQLCYHHS	470
483	Qy	DQLFRNPHQALLHTA-NRPDEBCVGBGLACHQLCARGHCWGPGTQCWNGSQFLRGQ	541
471	Db	TKVLRGTEBERLDIKHNRPRDCVAEGKVCDPCLCSSGCGWPGGQCLSCRNYSRGVCV	530
542	Qy	EECHVLOGLPREYVNAHCLPCHCECOPQNGSVTCFGEADOCVACHYKDPPECVACRP	601
531	Db	THCNFLNGEPREFEAECFSCHECOPMEGTATCNGSGSDTCAQCAHFRDGPCHVSSCP	590
602	Qy	SGVKPDLSPYMPKFPDDEEGACQPCPNCHTSCVDLDDKGC	653
591	Db	HGVLG--AKGPIYKVDQNECRFCHENTCGCKGPELQDCLGQTLVLIGKTHLTMA	648
654	Qy	IQYIKANSKFITITELKRQOKIR-KYTMRRLLQETELVPELTPSGAMPNQAQMRILKET	712
649	Db	IAGLWVIFMMLGGTFLYWRGRIQNKRAMRYLERGESIEPLDPS-EKANKVLARI	707
713	Qy	ELRKVKVLGSGAFVTYKGIWIPQENVKIPVAIKVIRENTSPKANKEILDVAVMAGV	772
708	Db	ELRKLKVLGSGVGTVHKVGMWIPGESIKIPVCIKVIDKSGRGSFOAVTDHMLAIGSLD	767
773	Qy	SPVYSRLIGICLTSTVOLVTOLMPYGCLLDHVRENRRGLSGODLLNMCWIAKMSYLE	832
768	Db	HAHIVRLGLCPGSSQLVTQYLPGLSLLDHVRQHRGALGPOLLNNGVQIAKMYILEE	827
833	Qy	VRLVHRDLAARNVLVSPNHVKITDFGLARLLDIDETEHADGKVPKIMWALESILRRR	892
828	Db	HGMVHRNLAARNVLLKSPSQVQVADFGVADLLPDDKQLLYSEAKTPIKMALESIHFG	887
893	Qy	FTHOSDWSYGVTVWELMTFCAKYDGIIPAREIPDLLEKGERLPQPICTIDVYMWKVC	952
888	Db	YTHOSDWSYGVTVWELMTFCAEYAGYLAEAEPDLLEKGERLAQPOICTIDVYMWKVC	947
953	Qy	WMIDECRPFRELVSFBSRMARDPQRFVITQNEIDLGA--SPLDSTFYRSLLDDDMG	1009
948	Db	WMIDENIRPTEKELANFTRMARDPRLVLIKRES-GPGIAPGEPHGLTNKULEEVELE	1006
1010	Qy	DLVDAEYLVPOQGFCCPDAPGAGGMVHRRHSSTRSGGDDTLGLPE-SEEEAPRSP	1068
1007	Db	PELDLDLDEAED-----NLATTLGSLSLPVGTLNRRPGSQSL	1047
1069	Qy	LAPSEGAGSDVFDGDLGMAAKGLQSLPTH-DPSPLYRSDPTVPLP-----SETDGY	1121
1048	Db	LSPSSGY-MPMNQNLGESCQESAVSGSERCPRPVSLLH-----PMPRGCLASESEGH	1100
1122	Qy	VA-----PLTCSPOPE-----YVNPQDVRPOPSPRSGP-----	1150
1101	Db	VTGSEAELOEKVMCRSRSRSPRPGDSAYHSQRHSLLTPTVTLSPGLLEEDVNGVY	1160
1151	Qy	LPAARPAGATILERAKTLS-P-KNGV-----KDVFAFGGAVENPEYLTPOGGAPOPH	1203
1161	Db	MPDTHLKGTPSSREGTLSSVGLSSVLGTEEDED-----EEVEYNNRRRRHSP-PHP	1211
1204	Qy	PPAFSPAPDNLYYWD-----QDPPERGAPSPSTFKGTPTAENPEYL	1243
1212	Db	PPSSLEELGEYMDVSGDLSASLGSTQSCPLHPVPIIMPATGATPDEDEYIM	1263

RESULT 10

JC4387
Epidermal growth factor receptor homolog precursor - rat
N:Alternate names: ErbB3 protein; HER3 protein
C:Species: Rattus norvegicus (Norway rat)
C:Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C:Accession: JC4387
R:Hellayer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A:Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.

Db 699 ARIPEKTELRLKVLGSGVGTGVIHKGWIPEGESIKIPVICIKVIEDKSGROFOAVTDHM 758
Qy 766 YVMAGVSPYVSRLLIGICLTSTVOLVTQMPYGCLLDHRVNRGRGLSGQDILNMCQIAK 825
Db 759 LAVGSLDHAHIVRLILGLCPGSSQLQVLTQYPLUGSLLDHVKHQREITGLPQLILNMGVQIAK 818
Qy 826 GMSVLELDVRLVHRDLAARNVLKSPNPKITDFGLARLLDIDETEHADGKVPKIKMAL 885
Db 819 GMYLLEHSMVHRDLALNVLKSPQVQVADFGVADILPPDKQLLHSEAKTPIKMAL 878
Qy 886 ESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDV 945
Db 879 ESIHFGKTHQSDVMSYGVTVWELMTFGAEYAGLRLAEIPDLLEKGERLAQPOICTIDV 938
Qy 946 YMIWVKMWDSECRPRFELVSEFSRMDRPPQFVVIQNEQDLGFPASPLDSTFYRSLLED 1005
Db 939 YMWVKMWDENIRPTKELANEFTRWARDPPRYLVIKRAS-GPGTF--PAEPSPVLTT 995
Qy 1006 DDMGDLVDAEYLVPOQGFPCDPAPGAGGMVHRHRSSTSGSGGDLTLGLEPSEE--- 1062
Db 996 KEL-----QEAELEPEL-----DIDLLEAESEGLA 1021
Qy 1063 -----EAPRSLAPSPSG-----AGSDVFDGLGMAAGKGLQSLPT 1097
Db 1022 TSLGALSILPTGLTRPRGSSLLSPSGGYMNMQSSLGACLOSALVGGREQFSRPLSL 1081
Qy 1098 HDPSPLOQYSEDPVPLPSETDGYV---APL-----TC-----SQPR-----YVNPQ 1137
Db 1082 H-PIPRGR-----PASESEGHVGTGSEAELEQKVSVCRSRSRSPRPRGDSAYHSQR 1133
Qy 1138 DVQPQSPSPRGP-----LPAARPAAGATLERAKTLSP-GKNGV-----KDVF 1179
Db 1134 HSLLTPTVPLPSPGLEEBDNGYVMPDTHLRGASSRREGTLSSVGLSGSLVLTGEBEDED-- 1191
Qy 1180 AFGGAVENPEYLTPOGGAPOPHR 1204
Db 1192 -----EYEYNNRRKRGSP-PRPP 1209
RESULT 11
TVFVLV
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
N;Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C;Species: avian leukosis virus, ALV
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C;Accession: B00643; A00643
R;Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A;Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: B00643
A;Molecule type: mRNA
A;Residues: 1-698 <NTIL>
A;Cross-references: GB:M10066; GB:M13881; NID:g211749; PIDN:AAA48763.1; PID:g211750
A;Note: in Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gal
C;Comment: This protein is synthesized as a gag-env-erbB protein.
C;Genetics:
A;Gene: gag-env-erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
F;1-6/Product: gag protein (fragment) #status predicted <GAG>
F;7-59/Product: env protein (fragment) #status predicted <ENV>
F;60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F;194-459/Domain: protein kinase homology <KIN>
F;202-210/Region: protein kinase ATP-binding motif
F;223/Active site: Lys #status predicted

Query Match 25.4%; Score 1723.5; DB 1; Length 698;
Best Local Similarity 51.5%; Pred. No. 3.6e-64;
Matches 368; Conservative 74; Mismatches 144; Indels 129; Gaps 17;
Qy 578 GPEADQCVACHYKDPFPCVARCPSGVKPDLSYMPIWKPPEEGACQPCPNCTHSCVDL 637

Db 60 GP--DHCMCKCAHFDIGHPCVKACPAAGVLGENDTL-VWKYADANAVCOLCHPNCTRCKGCP 116
Qy 638 DDKGCPEAEQASPLTSIQYIKANSKF-----IGITELKRRQKQIRKYTMRRLLIQETELV 691
Db 117 GLEGCP---NGSKTPSIAAGVVGSLLCVVVGLGILYLRHHIVRKRTRLRLQLQERELV 173
Qy 692 EPLTPSGAMINQAMRILKETELRKVKVGLSGAGFTVYKGIWIIPDGENVKIPVAIKVLE 751
Db 174 EPLTPSGEAPNQALHRLIKETEFKVKVGLSGAGFTVYKGLWIPGEKVKIPVAIKLE 233
Qy 752 NTSPKANKETLDRAYMAGVGSPPVSRLLGICLTSTVOLVTQMPYGCLLDHRVNRGR 811
Db 234 ATSPKANKETLDRAYMVASVDNPHVCKLLGICLTSTVOLVTQMPYGCLLDIREHNDNI 293
Qy 812 GSQDLNMCQIAKMSYLEDVRLVHRDLAARNVLKSPNPKITDFGLARLLDIDETEY 871
Db 294 GSQYLLNWCYQIAKGMNVLBEERLVRDLAARNVLKTPQHVKITDFGLAKLLGADEKEY 353
Qy 872 HADGGKVPKIKMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEK 931
Db 354 HADGGKVPKIKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSVLEK 413
Qy 932 GERLPOPICTIDVYMIWVKMWDSECRPRFELVSEFSRMDRPPQFVVIQ-NEDLGP 990
Db 414 GERLPOPICTIDVYMIWVKMWDADSRPKFRELIAEFSKMDRPPRYLVILQGERMHL 473
Qy 991 ASPLDSTFYRSLLEDMDGDLVDAEYLVPOQGFPCDPAPGAGGMVHRHRSSTSGG 1050
Db 474 PSTDSKFYRLMEEDMEDIVDAEYLVPHQGF-----NPFST----- 513
Qy 1051 GDLTLGLEPSEEBAPRSL-----APSEGAGSVDFDGLGMAAGKGLQSLTHTDPSPLQR 1105
Db 514 -----SRTPLLSSLSATSNNSATNCID-----RNGQGHVPVREDSFVQR 551
Qy 1106 YSEDPVPLPSET--DGYVAPLTCSPQPEYVQPDVVRPQPPSPREGPLPAARPAATLER 1163
Db 552 YSSDPTGNFLUESIDDGFL-----PAPEYVQ--LMPKKPS----- 585
Qy 1164 AKTLSPGKNG/VKDVF-----AFGGAVENPEYLTPOGGAAPQPHPPAFS 1208
Db 586 ---TAMVQNIYNNISITAIKLPMSRYQNSHSTAVDNEPYL-----NTNQSPAK 634
Qy 1209 PAFDNLVYWDQ-----DPPE-----RGAPPSTFKGTPTAENPEYLGLDVP 1248
Db 635 TVFESSPYMTQSGNHQINLDNPDYQDFLFPNETKPNGLLKVPAEAENPEYLRVAAP 689
RESULT 12
TVYUW
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)
C;Species: avian erythroblastosis virus
C;Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C;Accession: A00644; A38022
R;Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A;Title: The erbB gene of avian erythroblastosis virus is a member of the src gene family
A;Reference number: A00644; MUID:84026539; PMID:6313229
A;Accession: A00644
A;Molecule type: DNA
A;Residues: 1-604 <YAM>
A;Cross-references: GB:K01216; NID:g209676; PIDN:AAA42400.1; PID:g209678
R;Debuire, B.; Henry, C.; Benaisa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Martin, J.
Science 224, 1456-1459, 1984
A;Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type of
A;Reference number: A38022; MUID:84223957; PMID:6328658
A;Accession: A38022
A;Molecule type: DNA
A;Residues: 1-28,'W',30-139,'F',141-145,'V',147-152 <DEB>
A;Cross-references: GB:K02006
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p

Db 655 QRECFQRHEC---NG---CTGPGADCKSRNFKLFDANETGYVNSTMFCNCKSKPLE 708
Qy 604 VK-PDLSYMPYKFPDEBACQPCPINCTHSCVLDKDCPAEQRASPLTSIQYIKANSK 662
Db 709 MRHVNQYTAGPY-----CAASP-----PRSSKITA--NLDVNM 742
Qy 663 FI-----GITEKRRQOKIRKYT--MRRLLQETELVEPLTPSGAMPNQAQM 706
Db 743 FIITGAVLVPTICILCVVTVYICRQKQAKKQAKETVMTWALSREDSEPLRPSNIGANLCKL 802
Qy 707 RILKETELRVKVLGSGAGFYVGIWIPDGENVKIPVAIKVIRENTSPKANKILDEAY 766
Db 803 RIVDAELRGVGLGMGAFGVYGVWVPEGENYKIPVAIKELKSTGAESSEFLEAY 862
Qy 767 VMAGVGSPPYRLLGICLTSTVQLVTLMPYGCCLLDHVRNRLGSDLLNLCWQIAKG 826
Db 863 IMASEEHVLLKLLAVCMSSOMLITQLMPLGCLLDVYRNRDKIGSKALLNWSQIAKG 922
Qy 827 MSYLEDVRLVHRDLAARNVLK---SPNHVKITDFGLARLLDIDETEYHADGGKVPKWM 883
Db 923 MSYLEERLVRHDLAARNVLVRLLAGEDH---DFGLAKLLSSDSNEYKAAGGKMPKWL 978
Qy 884 ALESILRRRPTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTI 943
Db 979 ALECIIRNVFTSKSDVWAFGTIWEILLTFQRPHEHIPAKDIPDLIEVGLKLEQPEICSL 1038
Qy 944 DVYIMVWKMWIDSECRPRFRELSEFSRMDARDPQRFVVIQNEDLG--PASPLDSTFYS 1001
Db 1039 DIYCTLLSCHWLDAAKPTFKQLITVFAEPARDGRVLAILGDKFRLPA-----YTS 1091
Qy 1002 LLED--DMGDLVDAEYLVPOQGFCCPDPAAGAGMVHRRSSSTRSGGDLTLGLE 1058
Db 1092 QDEKDLIRKLAPTGDSEATAKPDYILQKALGPS-----HRTDCT----- 1133
Qy 1059 PSEEEAP-----RSPLAPSEGASGVFDG---DLGMAAGKGLQSLPHTDPSLQRYSED 1109
Db 1134 ---DEMPKLNRYCKDPKNSKNSSTGDDERDSSAREVGVGNLR----- 1171
Qy 1110 PTVPLPSETDGYVAPLTCSPQEVYVNPQDVRPQPSREGPLPAARPAAGATLERAKTLSP 1169
Db 1172 --LDLPVDEDDYLMPT-QCPQPNNNNNN-----NPNQNNAAVGAAGYM----- 1214
Qy 1170 GKNGVVKDVFAGGAVENPEYL-----TPQGAAPQPH-----PPPAF 1207
Db 1215 -----DLIGVPVSDNPEYLLNAQTLGVGESPIQTIGIPVMGGPGTMEVKVPMGS 1267
Qy 1208 SP-AFDNLYWD 1218
Db 1268 EPTSSDHEYND 1279

RESULT 14

S35745
C;Species: protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
C;Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
C;Accession: S35745
R;Vennstrom, B.
Submitted to the EMBL Data Library, March 1993
A;Reference number: S35743
A;Accession: S35745
A;Molecule type: DNA
A;Residues: 1-544 <VEN>
C;Genetics:
C;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
F;135-400/Domain: protein kinase homology <KIN>
F;143-151/Region: protein kinase ATP-binding motif
F;170/Active site: Lys #status predicted

Query Match 23.6%; Score 1604; DB 2; Length 544;
Best Local Similarity 54.1%; Pred. No. 2.3e-59;
Matches 339; Conservative 64; Mismatches 128; Indels 96; Gaps 14;
Qy 578 GPEADQCVACAHYKDPFPFCVARCPSGVKPDLSYMPYKFPDEBACQPCPINCTHSCVDL 637
Db 1 GP--DHCMKCAHFDGHCVCAPAGVLGENDTL-VWKYADANAVCOLCHPNCNTRGCKGP 57
Qy 638 DDKCPAEQRAFSPLTSIQYIKANSK-----IGITELKRRQOKIRKYTMRRLLOETELV 691
Db 58 GLEGCP---NGSKTPSIAAGVVGGLLCLVVGIGLYLRRRHVVKRTLRLLQERELV 114
Qy 692 EPIPTSGAMPNQAQMILKETELRVKVLGSGAGFYVGIWIPDGENVKIPVAIKVIRE 751
Db 115 EPLTPSGEAPNQAHLRIKETEFKVKVLGFGAGFYVGIWIPDGENVKIPVAIKVIRE 174
Qy 752 NTSPEKANKETILDEAYVMAGVGSPPYRLLGICLTSTVQLVTLMPYGCCLLDHVRNRLG 811
Db 175 ATSPKANKETILDEAYVMASVDNPHVCRLLGICLTSTVQLVTLMPYGCCLLDYIREKDN 234
Qy 812 GSODLLNWCMIAGKMSYLEDEVLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEY 871
Db 235 GSQYLLNWCVCIAKGNWYLEERHVRDLAARNVLKTPQHVKITDFGLAKQLGADKEY 294
Qy 872 HADGKVPKIKMALESILRRRPTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEK 931
Db 295 HAEGKVPKIKMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPIAREIPDLLEK 354
Qy 932 GERLPPOPICTIDVYIMVWKMWIDSECRPRFRELSEFSRMDARDPQRFVVIQNEDLG 990
Db 355 GERLPPOPICTIDVYIMVWKMWIDSECRPRFRELSEFSRMDARDPQRFVVIQNEDLG 414
Qy 991 ASPLDSTFYRSLLEDDMGDLVDAEYLVPOQGFCCPDPAAGAGMVHRRSSSTRSGG 1050
Db 415 PSTDSKFTYLMNEEDMEDIVDAEYLVPOQGF-----NSPST----- 454
Qy 1051 GDLTLGLEPSEEBAPRSL-----APSEGASGVFDGDLGMAAGKGLQSLPHTDPSLQ 1105
Db 455 -----SRTPLLSLSATSNNSATNCIDRNGG-----H----- 481
Qy 1106 YSEDPTVPLPSETDGYVAPLTCSPQEVYVNPQDVRPQPSREGPLPAARPAAGAT-LERA 1164
Db 482 -----PVREDGFL-----PAPEYVQ--LMPKPKSTAMVQNIYVLSLTAKSL 524
Qy 1165 KTLSPGKNGVVDVFAFGGAVENPEYL 1191
Db 525 PIDSRVQN-----SHSTAVDNPEYL 544

RESULT 15

S00727
C;Species: kinase-related transforming protein (erbB) (EC 2.7.1.1) - avian erythroblastosis virus
C;Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
C;Accession: S00727
R;Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
Oncogene Res. 1, 265-278, 1987
A;Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mutant
A;Reference number: S00727; MUID:88217326; PMID:2897102
A;Accession: S00727
A;Molecule type: DNA
A;Residues: 1-545 <SCO>
A;Cross-references: EMBL:X06943
C;Genetics:
C;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; phosphotransferase
F;135-400/Domain: protein kinase homology <KIN>
F;143-151/Region: protein kinase ATP-binding motif

Query Match 23.5%; Score 1597; DB 2; Length 545;
Best Local Similarity 54.1%; Pred. No. 4.6e-59;
Matches 339; Conservative 63; Mismatches 129; Indels 96; Gaps 14;

```
Qy 578 GPEADQVACAHYKDPPECPVACRCPGVKPDLSYMPIMKFPDEGACOPCPINCTHSCVDL 637
Db 1 GP--DHCWKCAHFDGPHCVACFAGUGENDTL-VWKIADANAVCQLCHENCNTRGCKGP 57
Qy 638 DDGCPAEQASPLTSIOYIKANSKF-----IGITELKRQKQKIRKYTWRRLLQETELV 691
Db 58 GLEGCP---NGSKTPSIAAGVVGGLCLVAVVGLGIGLYLRRRHIVRKRTLRLRLQERELV 114
Qy 692 EPLTPSGAMPNOAQWRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRE 751
Db 115 EPLTPSGEAPNOAHLRIKETEFKKVKVLGFGAGFTVYKGLWIPGEKVTIPVAIKELRE 174
Qy 752 NTSKANKKEIIDEAVWAGVSPVYSRLGICLTSTVOLVTQLMFYGLDHDHVRNRL 811
Db 175 ATSPKANKKEIIDEAVWASVUNPHVCRLLGICLTSTVOLITQLMFYGCLLDYIREHKONI 234
Qy 812 GSQDLLNMCQIAKMSYLEDLVLRDLAARNVLKSPNHNKITDFGLARLLDIDETEV 871
Db 235 GSQYLLNWCQIAKGMNYLEERHLVHRDLAARNVLKTPQDVKITDFGLAKQLGADEKEY 294
Qy 872 HADGGKVPKMALESILRRFTQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEK 931
Db 295 HAEGGKVPKMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPIASEISSVLEK 354
Qy 932 GERLPQPPICTIDVTMIMVKCMWIDSECRPRFRELVSFEFSRMARDPORFVVIQ-NEDLGP 990
Db 355 GERLPQPPICTIDVTMIMVKCMWDASRPKRELIAEFSKWARDPPRYLVIOGDERMHL 414
Qy 991 ASPLDSTFYRSLLEDGDDGLVDAEYLVPOQGFPCPDPAFGAGGMVHRRHSSSTRSGG 1050
Db 415 PSPTDSKFYRTLMBEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
Qy 1051 GDLTGLPESEEAAPRSL-----APSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPLO 1105
Db 455 -----SRTPLLSLSATSNNSATNCIDRNG-----H----- 481
Qy 1106 YSEDPVPLPSETDGYVAPLTCSPQPEYVNOVDYRPOPPSPREGPLPAARPAGAT-LERA 1164
Db 482 -----PVREDFL-----PAPEYVNO-LMPKRPSTAMVQNIYNYISLTAISKL 524
Qy 1165 KTLSPGKNGVVKDVFAGGAVENPEYL 1191
Db 525 PMDSRYQN-----SHSTAVDNPEYL 544
```

Search completed: July 22, 2003, 09:27:20
Job time : 28.179 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:10:23 ; Search time 22.1435 Seconds
(without alignments)
5413.772 Million cell updates/sec

Title: SEQ4-653-675-12
Perfect score: 6776
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1247
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6602	97.4	1255	1 A24571	protein-tyrosine k
2	5809	85.7	1260	1 TVRTNU	protein-tyrosine k
3	5799.5	85.6	1254	2 I48161	p-185 precursor -
4	3066	45.2	1210	1 GQHUE	epidermal growth f
5	3032	44.7	1210	2 A53183	epidermal growth f
6	3011.5	44.4	1223	1 TVCHLV	epidermal growth f
7	2915.5	43.0	1308	2 A47253	epidermal growth f
8	2596	38.3	1166	1 S06142	protein-tyrosine k
9	2381.5	35.1	1342	2 A36223	kinase-related tra
10	2293.5	33.8	1339	2 JC4387	epidermal growth f
11	1683.5	24.8	698	1 TVFVLV	protein-tyrosine k
12	1620	23.9	604	1 TVYUHV	protein-tyrosine k
13	1612.5	23.8	1330	1 GQFFE	epidermal growth f
14	1564	23.1	544	2 S35745	protein-tyrosine k
15	1557	23.0	545	2 S00727	kinase-related tra
16	1540	22.7	540	2 B44776	protein-tyrosine k
17	1538	22.7	540	1 TVFVFB	protein-tyrosine k
18	1509	22.3	644	2 A36325	epidermal growth f
19	1287	19.0	1323	2 E88257	protein let-23 (im
20	1287	19.0	1374	2 S70712	protein-tyrosine k
21	1197	17.7	1369	2 S70713	protein-tyrosine k
22	1144	16.9	1717	1 A45558	epidermal growth f
23	1126	16.6	527	2 A42032	epidermal growth f
24	975.5	14.4	843	2 A27131	epidermal growth f
25	806.5	11.9	346	2 S13807	protein-tyrosine k
26	754.5	11.1	311	2 S13808	protein-tyrosine k
27	735	10.8	1363	2 T43220	insulin-like growth
28	704	10.4	1383	2 A36080	insulin receptor p
29	703.5	10.4	1372	2 A34157	insulin receptor p

ALIGNMENTS

RESULT 1

A24571

Protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N;Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein erbB

C;Species: Homo sapiens (man)

C;Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999

C;Accession: A24571; A25491; A44188; B4188; I59509; I57522

R;Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;

Nature 319, 230-234, 1986

A;Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth f

A;Reference number: A24571; MUID:86118663; PMID:3003577

A;Accession: A24571

A;Molecule type: mRNA

A;Residues: 1-1255 <YAM>

R;Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.

Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A;Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epiderm

A;Reference number: A25491; MUID:86016729; PMID:2995967

A;Accession: A25491

A;Molecule type: DNA

A;Residues: 737-1031 <SEM>

A;Cross-references: GB:M11767; NID:g182163; PIDN:AAA35808.1; PID:g553282

R;Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg, P.

Science 230, 1132-1139, 1985

A;Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromos

A;Reference number: A44188; MUID:86070181; PMID:2999974

A;Accession: A44188

A;Molecule type: DNA

A;Residues: 740-910 <COU1>

A;Cross-references: GB:M12036; NID:g183988; PIDN:AAA35978.1; PID:g183989

A;Accession: B44188

A;Molecule type: mRNA

A;Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>

A;Cross-references: GB:M11730; NID:g183986

R;King, C.R.; Kraus, M.H.; Aaronson, S.A.

Science 229, 974-976, 1985

A;Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A;Reference number: I59509; MUID:85272597; PMID:2992089

A;Accession: I59509

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 832-909 <REX>

A;Cross-references: GB:I29395; NID:g459807; PIDN:AAA35809.1; PID:g459808

R;Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987

A;Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptional

A;Reference number: I57622; MUID:87286898; PMID:3039351

A;Accession: I57622

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-191 <TAL>

A;Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C;Genetics: GDB:ERBB2; NGL; NEU; HER-2
A;Gene: GDB:ERBB2; NGL; NEU; HER-2
A;Cross-references: GDB:120613; OMIM:164870
A;Map position: 17q21.1-17q21.1
A;Introns: 25/1; 75/3; 147/1; 883/3
A;Note: the list of introns is incomplete
C;Function:
A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
inase
F;1-21/Domain: signal sequence #status predicted <SIG>
F;122-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F;122-653/Domain: extracellular #status predicted <EXT>
F;70-304/Domain: EGF receptor extracellular domain repeat <EEI>
F;395-605/Domain: EGF receptor extracellular domain repeat <EE2>
F;654-675/Domain: transmembrane #status predicted <TM>
F;676-1255/Domain: intracellular #status predicted <INT>
F;718-983/Domain: protein kinase homology <KIN>
F;728-734/Region: protein kinase ATP-binding motif
F;168,124,187,259,530,571,629/Binding site: carboxydrate (Asn) (covalent) #status predicted
F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;753/Active site: Lys #status predicted
F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 97.4%; Score 6602; DB 1; Length 1255;
Best Local Similarity 97.2%; Pred. No. 5.3e-265;
Matches 1220; Conservative 9; Mismatches 16; Indels 8; Gaps 3;

Qy 1 MELAALCWGULLALLPPGAASVQCTGDMKRLRASPETHDMLRHLHYGCGVQGNL 60
Db 1 MELAALCWGULLALLPPGAASVQCTGDMKRLRASPETHDMLRHLHYGCGVQGNL 60
Qy 61 ELTYLPTNASLSFLQIDQEVQGVYLIHNNQVQVQLRQLRVGRTQLFEDNYALAVLNG 120
Db 61 ELTYLPTNASLSFLQIDQEVQGVYLIHNNQVQVQLRQLRVGRTQLFEDNYALAVLNG 120
Qy 121 DPLNNTTPTVGASPGGLRELQRLSRLTEILKGVLIQVLPOLCYQDTILWKDIFHNQOLA 180
Db 121 DPLNNTTPTVGASPGGLRELQRLSRLTEILKGVLIQVLPOLCYQDTILWKDIFHNQOLA 180
Qy 181 LTLIDTNSRACHPCSPCKSRGWSESDCOSLTRVTCAGGCARCKGLPTCCHEQC 240
Db 181 LTLIDTNSRACHPCSPCKSRGWSESDCOSLTRVTCAGGCARCKGLPTCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESNPNPEGRTYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESNPNPEGRTYTFGASCVTACP 300
Qy 301 YNLTSTVGSCITLVCLNHNQVATBDGTQRCCKSPCARVCYGLMGQYIKANSKFIQIT 360
Db 301 YNLTSTVGSCITLVCLNHNQVATBDGTQRCCKSPCARVCYGLMGQYIKANSKFIQIT 360
Qy 361 ELEFAGCKKIFGSLAFLESFEDGDPASNTAPLOEQLOVFLEITCYLYISAWPDSLP 420
Db 361 IOEFAGCKKIFGSLAFLESFEDGDPASNTAPLOEQLOVFLEITCYLYISAWPDSLP 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWGLRSLRELGSGLALHNNHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWGLRSLRELGSGLALHNNHLCFVHTV 480
Qy 481 PWDQFRNPHOALLHTANRPEDECVGEGGLACHQLCARGHCWGPQTQCVCNSQFLRGQEC 540
Db 481 PWDQFRNPHOALLHTANRPEDECVGEGGLACHQLCARGHCWGPQTQCVCNSQFLRGQEC 540
Qy 541 VEEECVLOGLPREYNARHCLPCHPECOPOGNSVTCFGEADOCVACAHYKDPPECVARC 600
Db 541 VEEECVLOGLPREYNARHCLPCHPECOPOGNSVTCFGEADOCVACAHYKDPPECVARC 600
Qy 601 PSGVKPDLSPYMPKPFDEBEGACQPCPINCSTHSCVDLDDKGCAPAEQASPLTISAVVG 660
Db 601 PSGVKPDLSPYMPKPFDEBEGACQPCPINCSTHSCVDLDDKGCAPAEQASPLTISAVVG 660

Db 601 PSGVKPDLSPYMPKPFDEBEGACQPCPINCSTHSCVDLDDKGCAPAEQASPLTISAVVG 660
Qy 661 ILLVVLGVVFGILJ-----QYIKANS--KFIGITEL--PLTPSGAMPNQAMRILKETEL 712
Db 661 ILLVVLGVVFGILJKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 713 RKVKVLGSGAGFTGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVNAVVGSP 772
Db 721 RKVKVLGSGAGFTGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVNAVVGSP 780
Qy 773 YVSRLLGICLTSTVQLVQLMPLVYGLLDHVRNRLGRLGSDLLNWCQIAKMSYLEDV 832
Db 781 YVSRLLGICLTSTVQLVQLMPLVYGLLDHVRNRLGRLGSDLLNWCQIAKMSYLEDV 840
Qy 833 LVHRDLAARNVLKSPNHNKIDTDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 892
Db 841 LVHRDLAARNVLKSPNHNKIDTDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900
Qy 893 HQSDVWSYGVTVWELMTFGAKPYDGIPIREIPDLLEKGERLPQPICTIDVYMIWVKCW 952
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIPIREIPDLLEKGERLPQPICTIDVYMIWVKCW 960
Qy 953 IDSECRPRFRELVSFSEFARMARDPQRFVVIQNEDELGPASPLDSTFYRSLLEDDDDMGDLVDA 1012
Db 961 IDSECRPRFRELVSFSEFARMARDPQRFVVIQNEDELGPASPLDSTFYRSLLEDDDDMGDLVDA 1020
Qy 1013 EYILVPOQGFCDPPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1072
Db 1021 EYILVPOQGFCDPPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Qy 1073 AGSDVDFGDLGMGAAGLQSLPHTDPSPLORYSDPTVPLPSETDGVVAPLTCSPQPEYV 1132
Db 1081 AGSDVDFGDLGMGAAGLQSLPHTDPSPLORYSDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Qy 1133 NQDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPKXGVVYKVFAGGAVENPEYLTQP 1192
Db 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPKXGVVYKVFAGGAVENPEYLTQP 1200
Qy 1193 GGAAPQHPHPPAPAFDNLVYWDQDPPERGAPSTFKGTPTAENPEYGLDV 1247
Db 1201 GGAAPQHPHPPAPAFDNLVYWDQDPPERGAPSTFKGTPTAENPEYGLDV 1255

RESULT 2
TVRNU
protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
C;Accession: A24562; A61204
R;Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A;Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.
A;Reference number: A24562; MUID:86118662; PMID:3945311
A;Accession: A24562
A;Molecule type: mRNA
A;Residues: 1-1260 <BAR>
A;Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746
R;Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smath, R.A.; Cohen, Carcinogenesis 12, 1975-1978, 1991
A;Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no m 2-thiazolyl]formamide or N-methyl-N-nitrosourea.
A;Reference number: A61204; MUID:92035293; PMID:1682063
A;Accession: A61204
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 637-663,'V',665-702 <MAS>
A;Note: authors translated the codon GCA for residue 25 as Val
C;Genetics:
A;Gene: neu
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

F:658-680/Domain: transmembrane #status predicted <TM>
F:723-988/Domain: protein kinase homology <kin>
F:731-739/Region: protein kinase ATP-binding motif
F:71,191,263,535,576,634/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:691/Binding site: phosphate (Thr) (covalent) #status predicted
F:758/Active site: Lys #status predicted
F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

	Query Match	85.7%	Score 5809;	DB 1;	Length 1260;
	Best Local Similarity	85.6%;	Pred. No. 2.5e-232;		
	Matches 1076;	Conservative 55;	Mismatches 116;	Indels 10;	Gaps 5;
Qy	1	MELAAALCRWGLLLALLPPGAAS	TQVCTGTMKRLPASBETHLDMRLHLYQGQGVVQGNL	60	
Db	4	MELAAACRWGFLALLPPGIAGTQVCTGTMKRLPASBETHLDMRLHLYQGQGVVQGNL	63		
Qy	61	ELTYLPTNASLSFLQDIQSVQGVLI	TAHQVQVPLQRLIRVGTOLFEDNYALAVLDNG	120	
Db	64	ELTYVPAVASLSFLQDIQSVQGM	LIAHQVKKVPLQRLIRVGTOLFEDKYLAVLDNR	123	
Qy	121	DPLNNTTPTV	TGASPGGRELQRLSLEILKGGVLI	QIRNPQLCVQDTILWKDIFHKNNQL	179
Db	124	DPQDNVAASTPGRTPEGLRELQRLSLEILKGGVLI	RGNPQLCVQDMVLWKDIFRKNQL	183	
Qy	180	ALTLDITNRSRACHPCSPCKSGR	CSWGSESDCQSLTRTVCAAGCARCKGFLPTDCHEQ	239	
Db	184	APVDITNRSRACPPCAPACKONHCWGESPEDCQILTGTICTSGCARCKGRLPTDCHEQ	243		
Qy	240	CAAGCTGPKHSDCLACLFHNHSGI	CELHCPALVYNTDTFESMNPBEGRYTFGASCVTAC	299	
Db	244	CAAGCTGPKHSDCLACLFHNHSGI	CELHCPALVYNTDTFESMNPBEGRYTFGASCVTTC	303	
Qy	300	PYNYLSTDVGSC	TLVCPILHNQVTAEDGTQRCCKSPCARVCYGLCMQVIAKNSKFIGI	359	
Db	304	PYNYLSTVGSC	TLVCPNNQVTAEDGTQRCCKSPCARVCYGLMEHLRGARITSD	363	
Qy	360	TELEFAGCKKIFGSLAFLPES	FDGDPASNTAPLOPEQLQVFLEBITGYLIYSAMPDSL	419	
Db	364	NVQEFDGCKKIFGSLAFLPES	FDGDPSSGIAPLRPEQLQVFLEBITGYLIYSAMPDSL	423	
Qy	420	PDLVSFQNLQVIRGILHNGAYS	LTLOGILGHSWLGHSRLSRELGSGLALIHNTLHCFVHT	479	
Db	424	RDLVSFQNLRIIRGILHDGAYS	LTLOGILGHSRLSRELGSGLALIHNRNHLHCFVHT	483	
Qy	480	VPDQLFRNP	HOALLHTANRPEDE-CVGEGLACHQLCARGCHGWGPGPTQVCNCSQFLRGQ	538	
Db	484	VPDQLFRNP	HOALLHSGNRPEDELCVSSGLVCNSICAHGCHGWGPGPTQVCNCSHFIRGQ	543	
Qy	539	ECVEECRVQLG	LPRYVNAHCLPCHPECOPQNGSVTFCPEADQCACAHYKDPPEVA	598	
Db	544	ECVEECRVWKG	LPREVYSDKRLPCHPECOPQNSSETCFGSEADCAACAHYKDSSCVA	603	
Qy	599	RCPSGVKPDLS	WMPYTWKPEDEGACOPDINCTHSCVDLDDKCPAEORASPLTISVAV	658	
Db	604	RCPSGVKPDLS	WMPYTWKYPDEEGICQPCPINCTHSCVDLDERGCPAEQASPTFFIATV	663	
Qy	659	VGILLVVLGVVFG	ILII---QVIKANS--KFIGITEL--PLTPSGAMPNQAQMRILKET	710	
Db	664	EGVLLFLILVVVG	ILIIKKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKET	723	
Qy	711	ELRKVKVLGSG	AFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVG	770	
Db	724	ELRKVKVLGSG	AFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVG	783	
Qy	771	SPYVSRLLGI	CLTSTVQLTQMPYGLLDHVRNRRGLSGDILLNWCQIAKMSYLE	830	
Db	784	SPYVSRLLGI	CLTSTVQLTQMPYGLLDHVRHRRGLSGDILLNWCQIAKMSYLE	843	
Qy	831	VRLVHRDLAARNV	LKSPNHVKITDFGLARLLDIDETEVHADGCKVPIIKWMALESILRRR	890	
Db	844	VRLVHRDLAARNV	LKSPNHVKITDFGLARLLDIDETEVHADGCKVPIIKWMALESILRRR	903	
Qy	891	FTHOSDVMSYGV	TWMLMTFGAKPDGIPAREIPDLLEKGERLPQPPICTIDVVMIMVKC	950	

A;Molecule type: mRNA
A;Residues: 969-971, 'K', 973-1115, 'D' <ES>
A;Cross-references: EMBL:Z12608
R;Heiseermann, G.J.; Gill, G.N.
J. Biol. Chem. 263, 13152-13158, 1988
A;Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
A;Reference number: A28941; MUID:88330814; PMID:3130233
A;Accession: A28941
A;Molecule type: protein
A;Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000; 1002-1009,
R;Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
submitted to the EMBL Data Library, April 1994
A;Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
A;Reference number: S45325
A;Accession: S45325
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-971, 'K', 973-1210 <VER>
A;Cross-references: EMBL:X78987; NID:G488030; PIDN:CAA55587.1; PID:G488831
R;Paria, B.C.; Das, S.K.; Andrews, G.K.; Dev, S.K.
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
A;Title: Expression of the epidermal growth factor receptor gene is regulated in mouse b
A;Reference number: 149643; MUID:93126380; PMID:7678348
A;Accession: 149643
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 12-20, 22-132 <RES>
A;Cross-references: GB:L06864; NID:G193001; PIDN:AAA53029.1; PID:G567201
C;Genetics:
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop
F;1-24/Domain: signal sequence #status predicted <SIG>
F;648-670/Domain: transmembrane #status predicted <TM>
F;712-977/Domain: protein kinase homology <KIN>
F;720-728/Region: site: phosphate (Thr) (covalent) #status experimental
F;680, 695/Binding site: phosphate (Thr) (covalent) #status experimental
F;697, 1070, 1071/Binding site: phosphate (Ser) (covalent) #status experimental
F;993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
F;1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
F;1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 44.7%; Score 3032; DB 2; Length 1210;
Best Local Similarity 48.8%; Pred. No. 6, 3e-118;
Matches 621; Conservative 170; Mismatches 363; Indels 118; Gaps 26;
QY 11 LLLALLPPGAA--STQVCTGTDMLRLPASPETHDMLRLHYLQGCQVQVQGNLELYLPTN 68
DB 14 LLTALCAAGALEEKVKCGTSNRLTQLGTPEHFLSLQRMVNNCEVVLGNLEITYQVN 73
QY 69 ASLSFLQIOEQVGVYLAHNOVROVPLQRLRIVRGTOIFEDNYALAVLDNGDPLNNTTP 128
DB 74 YDLSFLKTIQEVAGVYLAHNTVERIPLENQIIRGNALYENTYALATLSN----- 124
QY 129 VTGASPGGLRELQRLSLTEILKGGVLIQKNPOLCYQDTILWKDI----FHKNNOLATLI 184
DB 125 -YGNRTGLRELPMNLQELLIGATVPSNNPLCNMDTIQWRDIQVNVFMSNMDL---- 180
QY 185 DTRNSRACHPCSPMKGSRGSESSDCQSLTRTVACGCA-RCKGPLPTCCCHQCAAG 243
DB 181 -QSPSPSCPKCDPSPGSCWGGGSENCQKLTIIQAQCSHRCGRSPSCCHNQCAAG 239
QY 244 CTGPKHSCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFEGASCVTACPNY 303
DB 240 CTGPRSDCLVCQKQFDEBATCKDTCPPLMLYNPTTYQMDVNPPEGKYSFGATCKKCPKNY 299
QY 304 LSTDVGSCTTLVCLPHNQEVATBDGTCQCEKSKPCARVCYGLGMQYIKANSKFIGITELE 363
DB 300 VYTHGSCVRACGPDYEV-EDGIRKCKKCDGPKVCNGIGICEFK-DTILSINATNIK 357
QY 364 -FAGCKITGSLAFIPESFGDGPASNTAPLOEQLOVFTLEITGYLYISAWPDSLPL 422
DB 358 *HPKYCTAIGDHLIPVAFKGSFTRTPPLDPRELEILKTKVEITGFLLIQAWPDNWDL 417

RESULT 6

TVCHLV

epidermal growth factor receptor precursor - chicken

N;Contains: protein-tyrosine kinase (BC 2.7.1.112) erbB

C;Species: Gallus gallus (chicken)

C;Date: 28-Feb-1996 #sequence_revision 05-May-1995 #text_change 04-Feb-2000

C;Accession: A27720; A00643

R;Lax, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennstr

Mol. Cell. Biol. 8, 1970-1978, 1988

A;Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou

A;Reference number: A27720; MUID:88261272; PMID:3260329

QY 423 SVFONLQVIRGILHNGAYSILTLQGLGTSWGLRSLRSLRELGLALIHNNTHLCFVHTVPW 482
DB 418 HAFENLEIIRGTRKQHQFSLAVVGLNITSGLRSLRSLRSLRSLRSLRSLRSLRSLRSLR 477
QY 483 DQLFRNPHQALLHTANPEDECVCVGEGLACHQLCARGHCWGFGPTQCVCNCSOFLRGQECVE 542
DB 478 KKLFGTNPQTKIWNRAEKDKAVNHVCNPLCSSEGCWGPEDPCVSCQNVSRGECVE 537
QY 543 ECRVLOGLPREYNARHCLPCHPCQONGSVTCFGEADOCVACAHYKDPFCVRCPS 602
DB 538 KCNILEGEPREFVENSEICQHPCLPQAMNITCTGRPDNCIQCAHYIDPHVCVKTCPA 597
QY 603 GVKEDLSVMPITWPKPEDEGCQPCINCTHSCVDLDDKGCPCAPRASPLTSIVSAVVGIL 662
DB 598 GIMGENNTL-VKTYADANNVCHLCHANCTYGCAGPGLQGCVEWPSGPKIPSIATIGVGL 656
QY 663 L---VVVLGVVFGILIQIY---KANSKFIGITEL---PLTPSGAMPNQAORILKETEELRK 714
DB 657 LFIIVVALGICLFMRRRHIVRKRTRLRLLOERLEVEPLTPSGEAPNQAHLRLKETEPFK 716
QY 715 VKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKIILDEAYVMAGVSPYV 774
DB 717 IKVLGSGAFGTYYKGLWIPGEKVKIPVAIKELREATSPKANKIILDEAYVMASVDNPHV 776
QY 775 SRLIGICLTSTVOLAVTOLMPYGCLLDHVRENKRGSLGSDLLNWCQIAKGMVSYLEDVRLV 834
DB 777 CRLIGICLTSTVOLAVTOLMPYGCLLDHVRENKRGSLGSDLLNWCQIAKGMVSYLEDVRLV 836
QY 835 HRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRRPTHQ 894
DB 837 HRDLAARNVLKTPQHVKITDFGLAKLGAEEKEYHAEGKVPKIMMALESILHRIYTHQ 896
QY 895 SDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKCMID 954
DB 897 SDVWSYGVTVWELMTFGSKPYDGIIPASDISILEKGERLPPOPICTIDVYIMVKCMID 956
QY 955 SECPREELVSEFMRARDPQRFVLIQ-NEDLGPASPLDSTFVRSILLEDDMDGLVDAAE 1013
DB 957 ADSRPFRELLFESQARDPQRLVYQGDHMLPSTFNSFNFRALMDEDMEDVDVAD 1016
QY 1014 EYLVPOQGFPCDPAPGAGGVVHRRSSSTRSGGDLTLGLEPSEEEAPRSPLAPSGA 1073
DB 1017 EYLVPOQGF-----NSPST-----SRTPLLSLSA 1042
QY 1074 GSDVFDGLGMAKGLQSLTPHDPSPLORYSEDTVPLPSET--DGVPALTCSPQPEY 1131
DB 1043 TSN-----NSTVACINRNGSCRVKEDAFQRYSSDPTGAVTEDNIDDAFL-----PVPEY 1092
QY 1132 VNQPDVRPQPSREGPLPAARPAATLAKTLSPGKNGVVKDVFAGGAVENPEYL-T 1190
DB 1093 VNQ-SVPRPAGSVQNVYHNPOLHP-----APGRDLHYQN--PHSNAGVPEYLN 1141
QY 1191 PQGAAPQPHPPPPAFSPAFNLYYWQ-----DP-----PERGAPPTFKGTPT 1234
DB 1142 AQ-----PTCLSSGFSNPAKMIQKSHQMSLDNPDYQDFFPKETKNGIFKG-PT 1191
QY 1235 AENPEYLGIDVP 1246
DB 1192 AENAEYLRVAPP 1203

A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
A:Cross-references: GB:M20386
R:Nielsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: A00643
A:Molecule type: mRNA
A:Residues: 585-1223 <NIL>
A:Cross-references: GB:M10066
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
specific protein kinase
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F:31-654/Domain: extracellular #status predicted <EXT>
F:31-307/Domain: EGF receptor extracellular domain repeat <EE1>
F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F:655-677/Domain: transmembrane #status predicted <TM>
F:678-1223/Domain: intracellular #status predicted <INT>
F:719-984/Domain: protein kinase homology <KIN>
F:727-735/Region: protein kinase ATP-binding motif
F:136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #
F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:754/Active site: Lys #status predicted
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 44.4%; Score 3011.5; DB 1; Length 1223;
Best Local Similarity 47.8%; Pred. No. 4.5e-117;
Matches 620; Conservative 175; Mismatches 349; Indels 153; Gaps 29;

Qy 8 RWGLLLALLPGAA-----STVCTGTDMLRLPASPETHLMDLRHLHYQCCVQVGNLE 61
Db 13 RGAAVLVLLGLGVALCSAVEKKVCQGTNNKLTQLGHVEDHFTSLQRMYNCEVVLNLE 72

Qy 62 LTYLPTNASLFLQDIQVQVGLIAHQVQVPLRLRIVRGTLQFEDNVALAVLNGD 121
Db 73 ITYEHNRLTLFTKIQVAGVLLALNMVDVPLENQLIIRGNVLYDNSFALAVLSNH 132

Qy 122 PLNNTPTVTGASPGSLRLQLSLTEILKGLVLIQRPOLCQVDTILWKDIFHKNNOLAL 181
Db 133 -MNKTQ-----GLRELPMKRLSEILNGVKISNNEKLCNMOTLVNDIIDSRLK-PL 182

Qy 182 TLID-TNRSRACHPCSPCKGSRGWSESDCSQSLTRTVACGGCA-RCKGPLPTDCCHQ 239
Db 183 TVLDPASNLSCPKHPNCTEDHCWGAGEQNCQTLTKVICAQCSGRCGRKVPSPDCCHQ 242

Qy 240 CAAGCTGPKHSDCLACLPNHSGLCELHCPALVTYNTDTFESMPNPEGRYTFGASCYTAC 299
Db 243 CAAGCTGPRESDCLACRFRDATTCKTCCPLVLYNPTYQMDVNPCKYSFGATCVREC 302

Qy 300 PYNLYSTDVGSCTVLCPLHNOEVTAEDGTORCEKSKPCARVCYGLGQYIKANSKFTGI 359
Db 303 PHNVVTVDHGVSRCNTDTEYV-EENGVRCKCKDGLCSKVCNGIGELKGLIS-INA 360

Qy 360 TELE-FAGCKKIFGLAPLPSFGDPDASNTAPLOPEOLQVFTLEETGYLYTSAMPDS 418
Db 361 TNIDSFKNCTKINGDVSILPVAFLGDAFTKTLPLDPKKLDVFTVKEISGELLQAWPDN 420

Qy 419 LPDLVSFONLQVIRGLIHNGAYSILTLQGLISWGLSLRSLGSLALIHNTLFCVH 478
Db 421 ATDLVAFENLEIRGTRKQGOYSLAVNLKIQSLGLRSLKEISDGDITAMKNKLCYAD 480

Qy 479 TVPMDQLFRNPQALLHTANPECEVGEGLACHOLCARGHCWGPPTQCVNCSQFLRGQ 538
Db 481 TMNWRSLFATOSQTKIILQNRKNDCTADRHVCDPLSDVCWGPFGPHCFSCFFSRQK 540

Qy 539 ECVEECRVLOGLPREYVNAHCLPCHPECPQNG---SVTCFGEADQCVACAHYKDPFF 595

Db 541 ECVKQCNILQGEPRFERDSKCLPCHSECLVQNSTAYNTTCSGPGPHCMKCAHFIDGPH 600
Qy 596 CVARCPGSKVPDLSYMPIWKFPPDEBEGACQCPNCTHSCVDLDDKGCPCAEQASPLTSIV 655
Db 601 CVKACPAVLGENDTL-VKMYADANAVCOLCHPNCTRGCKPGLEGCP---NGSKTSPSIA 656
Qy 656 SAVV-GILLVVVLGVVFGIIL--QYI---KANSKFIIGTEL--PLTPSGAMPNQAQWIRL 707
Db 657 AGVVGGLLCLVWGLGIGLYLRRRHIVRKRTLRLQLQERLVEPLTPSGEAPNQAHLRL 716
Qy 708 KETELRKVKVLSSGAGCTVYKGIWIDGENVKIPVAIKVLRENTSPKANKEILDEAYMA 767
Db 717 KETEFKKVKVLSSGAGCTVYKGIWIDGENVKIPVAIKVLRENTSPKANKEILDEAYMA 776
Qy 768 GVSGPYVSRLLIGLCTSTVOLVQLMPYGCGLLDHVRNRLGSLQDGLNNCMQIAKMSY 827
Db 777 SVDPNPHVCRLLIGLCTSTVOLVQLMPYGCGLLDYIREHKONIGSQYLLNMCVQIAKMN 836
Qy 828 LEDVRLVHRDLAARNVLVSPNHVKITDFGLARLIDIDETEHADGGKVPKMMALLESIL 887
Db 837 LEERLVRDLAARNVLVKTQHVKITDFGLAKLGADEKEYHAEGGKVPKMMALLESIL 896
Qy 888 RRPRTHQSDVMSYGVTVWELMTFGAKYDGIPIPAEIPDLLEKGBRLPOPPITCTIDVYIM 947
Db 897 HRIYTHQSDVMSYGVTVWELMTFGSKFYDGIPIPAEISSVLEKGBRLPOPPITCTIDVYIM 956
Qy 948 VKCMIDSECRPRFRELVSFSSRMARDPQRFVVIQ-NEDLGPASPLDSTFYVRSLEDDDM 1006
Db 957 VKCMIDADSRRKRELIAFSSKWARDPPRYLVIOQDERMHLPSPTSKYRTLMREEDM 1016
Qy 1007 GDLVDAEYLVPOQGFPCPDPAFCAGGMVHRRSSSTRSGGDLTLGLPESEEPSP 1066
Db 1017 EDIVDADEYLVPHQGF-----NSPST-----SRTP 1042
Qy 1067 L-----APSGAGSDVFDGDLGMAAGLQSLPHTDPSLQRYSEDDTVLPST--DGY 1119
Db 1043 LLSLSLATSNNATNCID-----RNGQGHVPREDSFVQRYSSDPTGNFLESIDDG 1094
Qy 1120 VAPLTCSPQEPQNPQVPPSPREGPLPAARPAAGATLERAKTLPSPGKGVVVDYF-- 1177
Db 1095 L-----PAPEYVQ--LMPKKFS-----TAMVQNIYNNISLT 1125
Qy 1178 -----AFGAVENPEYLTPOGGAAPQHPPPAFSPAFDNLTYWDQ----- 1217
Db 1126 AISKLPMSRYQNSHSTAVDNPEYL-----NTNQSLAKTVFESSPYWIOSGNHOIN 1177
Qy 1218 -DPPE-----RGAPSTFKGTPTAENPEYLGLDVP 1246
Db 1178 LDNPDYQODFLPNETKPNGLLKVPAENPEYLRVAAP 1214

RESULT 7
A47253
epidermal growth factor receptor, HER4 - human
C:Species: Homo sapiens (man)
C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C:Accession: A47253
R:Piowman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; Ne
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A:Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal
A:Reference number: A47253; MUID:93189574; PMID:8383326
A:Accession: A47253
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1308 <PLO>
A:Cross-references: GB:L07868; NID:G337359; PIDN:AAB59446.1; PID:G337360
A>Note: sequence extracted from NCBI backbone (NCBIP:126842)
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor
F:716-981/Domain: protein kinase homology <KIN>
F:724-732/Region: protein kinase ATP-binding motif

Db 123 YQK-NPSSP--DYYQVGLKQLQLSNLTELISGGVVKVSHNPLNLCNVETINWMDIVDKTNP 179
Qy 180 ALTLDTNRSRACHPCSPCKMGKSGCWGESSEDQSLTRITVCAGGC-ARCKGPLPTDCCHE 238
Db 180 TWNLIPIHAFERQCKQCDHGVNGSCWAPGPGHCQKFTKLLCAEQCNRCRGPKPIDCCNE 239
Qy 239 OCNAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCUTA 298
Db 240 HCAGCTGPRATDCLACRDFNDGCKDTPPKIYDIIVSHQVNDPNIKYTFGAAVCYKE 299
Qy 299 CPYNLSLTDVSGCTLVCLPHNQEVTAEDGTORCEKSPKPCVVCGLGMQVQIKANSKFIG 358
Db 300 CFSNVVTE-GACVRSKSCAGMLEVD-ENGKSKCPDGVCPKVCUDGIGIGSL-SNTIAVN 356
Qy 359 ITEL-ERAGCKKIFKISLAFLPESFDGDPASNTAPIQEQLOVFETLEITGYLISAWPD 417
Db 357 STNIRSFNCTKINGDIIILNRNSPEGDPHYKIGTMDPEHLNLTITVKEITGYLVIMWPE 416
Qy 418 SLPDLSVFQNLQVIRGRILHNGAYS-LTLQGLGISWGLRLSRLRELGLALIHNTHLCP 476
Db 417 NMTSLSVFQNLLEIIRGRTTFRSGFSFVVVQVRHLQWLGLSLKEVSAGNVILKNTLQLRY 476
Qy 477 VHTVPWDLFRPHOALLHTANRPEDECVGEGIACHQI-CARGHCWPGPCTOCVNCQFLR 536
Db 477 ANTNIRLFRSEDOSIYDART-----ENQTCNNECSDGCW-PGPTMCVSLHVD 528
Qy 537 QOECVECRVLQGLPREYNARHCLPCHPEQOPQNGSVTCFGEADQCACAHYKDDPPFC 596
Db 529 GGRVCASNLQGEPREAQVGRGVQCHQECIVQDLSITCYGPGPANCKSAHFQDGPQC 588
Qy 597 VARPSPGVKPLSYMPIWKFPDEBAGCQCPINCTHSCVDLDDKCPAEBQASPLTSIVS 656
Db 589 IPRCPHGILGDDTL-IWKYADKMGCCQCPCHQNCCTQGCSPGLSGCRGD-IVSHSLAVG 646
Qy 657 AVVGLLWVLGVVFGILIQIKANSK----FIGITEL--PLTPSGAMPNOAQMRILKE 709
Db 647 LVSGLLITIVALLILVLRRLRRRIKRTKICLQEKELVELPLTPSGAPNOAFILRIKE 706
Qy 710 TELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAKVIRENTSPKANKEIIDEAYVWAGV 769
Db 707 TEFKDRVLGSGAGFTVYKGLWNPGENIRIPVAKVIREATSPKVNGEVLDEAYVWASV 766
Qy 770 GSPVYSRLGLCLTSTVOLVTQLMPYGCILDHVRENRRGLSGDQLLNWCMIQAKMSYLE 829
Db 767 DHPHVCRLGLCLTSAVOLVTQLMPYGCILDVYRQHQRICQWLLANVCVQIAKGMNYLE 826
Qy 830 DVRLVHRDLAARNVLKSPNHVKITDFGLARLLIDETEHADGKGKPIKMALESILRR 889
Db 827 ERHLVHRDLAARNVLLKNPNHVKITDFGLSKLLTADEKBYQADGGKVPKMALESILQW 886
Qy 890 RFTHQSDVWSYGVTVWELMTFCAPYDGIIPAREIPDLLEKGERLPQPPICITDVYTMIVK 949
Db 887 TYTHQSDVWSYGVTVWELMTFGSKPYDGIIPAKETASVLENGERLPQPPICITIEVMIILK 946
Qy 950 CWMIDSECRPRELVSFBSRMARDPQRFVVIQNEEDLGPASPLDSTFYRSLLEDMDGDL 1009
Db 947 CWMIDPSRPRPRELVSFBSQWADPSRYLVIQ--NLPSLSDRLPSRLSSDD--DV 1001
Qy 1010 VDAEYLVPQOGFCPPDPAPGAGGVHRRHSSSTRSGGDLTLGLEPSEEAAPRSLAP 1069
Db 1002 VDAEYLLPLPKRI-----NRQGS-----EPCIP 1024
Qy 1070 SEGAGSDVDFDGLGMAKGLQSLTPHDPSPLOYSEDPV-PLPSETDGVVAPLTCSPQ 1128
Db 1025 PTGH-----PVRENSITLRNISDPTQNALEKLDLGH----- 1055
Qy 1129 PEYNQPDVRPOP-----PSPRE-----GELP-AARPAGATLERAKTILSPGRKNVVKD 1175
Db 1056 -EYVNPQGSSETSRSLSDIYNPNYEDLTDCWGPVSLSSQEAETNFSRPEYLNTQNSL--- 1111
Qy 1176 VFAFGGAVENPEYLTPOGGAAPQPHPPAPSPADNLYWQDPPERGAPSTFKGTPTA 1235
Db 1112 PLVSSGSMDDPDY---QAG-----YQAAF-----LPQTGALTGNWFLPAA 1149

Qy 1236 ENPEYLG 1242
Db 1150 ENLEYLG 1156

RESULT 9

A36223

kinese-related transforming protein (erbb3) (EC 2.7.1.1-) precursor - human
C:Species: Homo sapiens (man)
C:Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000

C:Accession: A36223; I59164

R:Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.

Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989

A:Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal g

A:Reference number: A36223; MUID:90083234; PMID:2687875

A:Accession: A36223

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1342 <KRA>

A:Cross-references: GB:M29366

R:Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, G.J

Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990

A:Title: Molecular cloning and expression of another epidermal growth factor receptor-re

A:Reference number: I59164; MUID:90311312; PMID:2164210

A:Accession: I59164

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>

A:Cross-references: GB:M34309; NID:G183990; PIDN:AAA35979.1; PID:G306841

C:Genetics:

A:Gene: GDB:ERBB3; HER3

A:Cross-references: GDB:119880; OMIM:190151

A:Map position: 12q13-12q13

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo

C:Keywords: ATP; phosphotransferase

F:707-972/Domain: protein kinase homology <KIN>

F:715-723/Region: protein kinase ATP-binding motif

Query Match 35.18; Score 2381.5; DB 2; Length 1342;
Best Local Similarity 40.18; Pred. No. 4.5e-91;
Matches 526; Conservative 197; Mismatches 450; Indels 139; Gaps 35;

Qy 10 GLLIALLPPGAA--STQVCTGTDMKRLRPLASPETHLDMRLHYQGCQVQVQNLBLYLPT 67
Db 11 GLUFSLARGSEVGNQAVCPFTLNGLSVTGDAENQYQTLKLYERCEVVMGNLBIVLTH 70
Qy 68 NASLSFLQDIQVOGYVLIHQNVRQVPLORLRIVRGTQLFEDNYALAVLDNGDPLNNTT 127
Db 71 NADLSFLQWIREVTGYVLVAMNBFSTLPLNLRVVRGTQVVDGKFAIVM----LNVNT 125
Qy 128 PVTGASPGGLREQLRSLRTEILKGGVLIQRNPOLCYQDTILWKDIFHKNQLALTLDTN 187
Db 126 ----NSSHALRQLRLTQLTEILSGGVYIERNDKLCHMDTIDWRDIVDRD---AEIVVKD 178
Qy 188 RSRACHPCSPMKXGSCWGESSEDQSLTRITVCAGGC-ARCKGPLPTDCHEQCAAGCTG 246
Db 179 NGRSCPCHEVCCKG-RCWGPGSEDCQTLTKTI-CAPQCNHCFGPNFNQCCCHDECAGCGSG 237
Qy 247 PKHSDDLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLST 306
Db 238 PQDTDCFACRHFNDGACVPRCPQPLVYNKLTQLEBNPHTKYQYGGVGVASCAPHNFV-V 296
Qy 307 DVGSCITLVCLPHNQEVTAEDGTORCEKSPKPCVVCGLGMQVQIKANSKF--IGITILE- 363
Db 297 DQTSQVRCACPPDKMEVD-KNGLKMCPCGGLCPKACEGTG-----SGSRFTQVTDSSNIDG 350
Qy 364 FAGCKKIFGSLAFLPESFDGDPASNTAPIQEQLOVFETLEITGYLISAWPDSLPLDS 423
Db 351 FVNCTKILGNLDFLITGLNGDPWHKIPALDPEKLVNFRVREITGYLNIOSWPPHMFNS 410
Qy 424 VFQNLQVIRGRILHNGAYS-LTLQGLGISWGLRLSRLRELGLALIHNTHLCLFVHTVPV 482

Db 411 VFSNLTITGRSLYNRGSFLLIMKLNVTSLGFRSLKEISAGRIYISANRQLCVHSLNW 470
Qy 483 DQLFRNPQALLHTA-NRPEDECVEGELACHQLCARGHCWGPGTQCVCNCSQFLRGQBCV 541
Db 471 TKVLRGPTTEERLDIKHNRPRDCVAEGKVCDFPLCSSGGCGWPGPGQCCLSCRNYSRGVCV 530
Qy 542 EECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPFCVARCP 601
Db 531 THCNFLNGEPREFAHEACFSCHPECPMEGTATCNGSGDSTCAQCAHFRDPHCVSSCP 590
Qy 602 SGVKPDLSPYMPFKPPEDEGACQPCPINCTHSCVDLDDKGCFAEORA-----SPLTSIYSA 657
Db 591 HGVLG--AKGPIYKYPDVQNECRPCHECHENTQCGKGPQLQCLGTLVLIGKTHLTALVT 648
Qy 658 VVGILLV-VVLGVF-----GILIQIKANSKFI--GITEPLPTPSGAMPNQAQWILKET 710
Db 649 TAGLVVFMMLGGTFLYWRGRRIQKRAMRYLERGESIEPLDPS--EKANKVLARIFKET 707
Qy 711 ELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVG 770
Db 708 ELRKVKVLGSGVGVTHGKWIPEGESIKIPVCIKVIEDKSGRQSFQAVTDHMLAIGSLD 767
Qy 771 SPYVSRLLGICLTSTVQLVTQLMPYGCCLLDHVRENRLGSGQDLNMCQIAKMSYLED 830
Db 768 HAHIVRLGLCPGSSQLVTQVPLGSLLDHVQRHGAIGALGPQLLNWGVQIAKMWYILEE 827
Qy 831 VRLVHRDLAARNVLKSNHVKITDFGLARLLDIDETEHADGGKVPKKNWALLESILRRR 890
Db 828 HGMVHRNLAARNVLKSPQVADFGVADLLPDDKQLLYSEAKTPKKNWALLESIHFGK 887
Qy 891 FTHQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPCTIDVYMWKVC 950
Db 888 YTHQSDVMSYGVYVWELMTFGAPYAGLRLEAEVPLLEKGERLAQPCICTIDVYMWKVC 947
Qy 951 WMIDSECRPRELSEFSRMARDQRFVVIQNEDLGPA-----SPLDSTFYRSLLEDDMG 1007
Db 948 WMIDENIRPTEKELANEFRMARDPPRYLVIKRES--GPIAPGPPHGLTNKLEVELE 1006
Qy 1008 DLVDAEYVLVQOQFCDDPAPGAGGMVHRRSSSTRSGGDLTLGLEP--SEEEAPRSP 1066
Db 1007 PELDLDLLEAED-----NLATTLGALSLSLPGVTLNRRPGSOSL 1047
Qy 1067 LAPSEGAGSDVFDGLGMAAKGLQSLPTH--PSPLQRYSEDPVPLP-----SETDGY 1119
Db 1048 LSPSSGY-MPMNQNLGSCQESAVSGSERCPRVSLH-----PMRPGCLASESSEGH 1100
Qy 1120 VA-----PLTCSPOPE-----YVNPQDVVRPQPPSPREGP----- 1148
Db 1101 VTGSEAELEQKVMCRSRSRSPRGRGDSAYHSQRHSLLTPTVPLSPGLEEDVNGVY 1160
Qy 1149 LPAARPAGATLERAKTILSP-GKNGV-----KDVFAFGGAVENPEYLTPOGGAAPOPHP 1201
Db 1161 MPDTHLKGTPSREGTSLSVGLSSVLGTEEDED-----EYEVYNNRRRHSP-PHP 1211
Qy 1202 PPARSPAFDNLVYWD-----QDPPERGAPPSTFKTPTAENPEYL 1241
Db 1212 FRPSSLEELGYEYMDVGDLSASLGSTOSCPHLHPVIMPTAGTTDPEDIETYM 1263

RESULT 10

JC4387
epidermal growth factor receptor homolog precursor - rat
N:Alternate names: ErbB3 protein; HER3 protein
C:Species: Rattus norvegicus (Norway rat)
C>Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C:Accession: JC4387
R:Hellyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A>Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.
A:Reference number: JC4387; MUID:96096535; PMID:8522190
A:Accession: JC4387
A:Molecule type: mRNA
A:Residues: 1-1339 <HEL>

A:Cross-references: GB:U29339; NID:g915389; PID:g915390
A:Experimental source: liver
A:Note: The authors translated the codon AAC for residue 369 as Thr and GTT for residue
C:Comment: This protein is a functional heregulin receptor that transduces signals to th
C:Genetics:
A:Gene: ErbB3
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo
C:Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1339/Product: epidermal growth factor homolog #status predicted <WAT>
F:640-659/Domain: transmembrane #status predicted <TM>
F:705-970/Domain: protein kinase homology <KIN>
F:713-721/Region: protein kinase ATP-binding motif
P:939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr) (

Query Match 33.8%; Score 2293.5; DB 2; Length 1339;
Best Local Similarity 40.3%; Pred. No. 1.9e-87;
Matches 517; Conservative 173; Mismatches 430; Indels 163; Gaps 36;

Qy 3 LAALCRWGLLLALLPPGAA---STQVCTGTDMLRLPASPEHLDMLRHLVQGCQVVOGN 59
Db 7 LQVLC-----FLSLARGSEMGNSQAVCPGLTNGLSVTGDDADNQYTLKLFKECEVVMGN 62
Qy 60 LELTYLPTNASLSPLQIQEYQGVYLIAHNQVRQPLQRLRIVRGTLQFEDNYALAVLDN 119
Db 63 LEIVLTGHNADLSFLQWIREVTAIVLVAMNEFSLPLNLRVVRGTQVYDGKFAIFVM-- 120
Qy 120 GDPNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYODTILMKDIFHKNQL 179
Db 121 ---LNYNT---NSSHALRQLKTLTEILSGGVYIEKNDKLCHMDTIDMRDVRVR--- 170
Qy 180 ALTIDITNRSPACHPCSPMKSGSEWCESSEDCQSLRTFTVCAGC--ARCXGPLDCCHE 238
Db 171 GAETIVKNGNANCPCHVEVCKG--RCWGPDPDCQILTKICAPQCNCRGCFPNPQCCHD 229
Qy 239 QCAAGCTGPKHSDCCLACLHFNHSGICELHCPALVYNTDTFESNPNPEGRYTFGASCVTA 298
Db 230 ECAGCGSGPQDTCFACRFNDSGACVPRCEPLVYNNKLTQLEPNHTKYQYGGVCVAS 289
Qy 299 CPYNYLSTDVGSCTVLCPLNHNEVTABDGTQRCBKSKPCARVCYL--GMQYIKANSKF 356
Db 290 CPHNFV--VDQTFVCRACPPDKMEVD--KHGLKMCPCGGLCPKACEGTGSGSRYQTVDSN 347
Qy 357 IGITELSPACKKTFGSLAFLPESFDGDPASNTAPLOEQLOVFEETELIYGLYVISAWP 416
Db 348 ID-----GFVNCIKLGNLDFLITGLNVDPWHKIPALDPEKLVNVRTREITGYLNIQSWP 403
Qy 417 DSLPDLVSFQNLQVIRGRIHNGAYS--LTQLGLGISWGLRLSLRELGLALIHHTHLC 475
Db 404 PHMNFVSFNLTTIGRSLYNRGFSLLIMKLNVTSLGFRSLKEISAGRVYISANQQLC 463
Qy 476 FVHTVPMQDLFRNPQALLHTA--NRPEDECVEGELACHQLCARGHCWGPGTQCVCNCSQF 534
Db 464 YHSLNLTLLRGPSEERLDIKYDRPLGECILAEGKVCDFPLCSSGGCGWPGAPQCCLSCRN 523
Qy 535 LRGEQCEVECEVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDP 594
Db 524 SREGVCVTHCNFLQGEPRFVHEAQCFPSCHPECLPMEGTSTYNGSGSDACARCAHFRDGP 583
Qy 595 FCVARCPGKVPDLSPYMPIWKFPDEEGACQPCPINCETHSC--VDLDDKGCPEAQASPLT 652
Db 584 HCVNSCPHGILG--AKGPIYKYPDAQNECRCHENCTQGCNGPELQDCLGAEVLMKSPH 641
Qy 653 SIVSAVGI--LLVVVLGVVP----GILIQIKANSKFI--GITEPLPTPSGAMPNQAQMR 705
Db 642 LVIAVTVGLAILMILGGSFLYWRGRRIQKRAMRYLERGESIEPLDPS--EKANKVLAR 700
Qy 706 ILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYV 765
Db 701 IFKETELRKVKVLGSGVGFVTHGKIWIPEGESIKIPVCIKVIEDKSGRQSFQAVTDHMLA 760
Qy 766 MAGVGSPPVSRLLGICLTSTVQLVTQLMPYGCCLLDHVRENRLGSGQDLNMCQIAKGM 825

Db 761 VGSLDHAIHVRLLGLCPGSSIQLVTVQYLPGLSLLDHVHQHRETLGPQLLLNWGVQIAKGM 820
QY 826 SYLEDVRLVHRDLAARNLVKSPNHNKVTDFGLARLIDIDETEHADGKVPKIKMALES 885
Db 821 YLEBHSVMVHRDLARNVWLSKSPSQVQVADFGVADLLPPDDKQLHSEAKTPIKMALES 880
QY 886 ILRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIARETIPDLLEKGERLPQPPICITIDVYM 945
Db 881 IHFGKYTHQSDVWSYGVTVWELMTFGABPYAGRLAEIPDLLEKGERLAQPOICTIDVYM 940
QY 946 IMVKCWMIDSECRPRFRELSEFSMARDPQRFVVIQNEIDGASPDLSTFYRSLLLEDD 1005
Db 941 VNVKCMWIDENIRPTFKELANEFTKWARDPPRYLVKRAK--GPGTP--PAEPPSVLTITKE 997
QY 1006 MGDVDAEYLVPOQGFCCPDPAFGAGGMVHRRHSSSTRSGGDLTLGLPSEB----- 1060
Db 998 L-----QEALEPEL-----DLDLLEABEGLATS 1023
QY 1061 -----EAPRSLAPSEG-----AGSDVDFDGLGMAAGLQSLPHD 1097
Db 1024 LGSALSPTGLTRPGSQSLSPSSGYMPNMNQSSLGAECLDSAVLGGREGQFSRPSILH- 1082
QY 1098 PSLQRYSEDTVPPLSPSTGTV-----APL-----TC-----SPOPE-----YVNOPDV 1137
Db 1093 PIPGR-----PASESEGHVTGSEAEALQKSVCKRSRSPRPGDSAYHSORHS 1135
QY 1138 RPQPPSPREGP-----LPAARPAAGATLERAKTLP--GKNGVV-----KDVPAF 1179
Db 1136 LLTPVTPPLSPGLEBEDNGYVMPDTHLGRASSREGTSSVGLSSVLGTREEDED----- 1191
QY 1180 GGAVENPRLTPOGGAAPQPHPP 1202
Db 1192 ----EVEYMNKRGRSP-PRPP 1209

RESULT 11
TVFVLV
N:Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C:Species: avian leukosis virus, ALV
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C:Accession: B00643; A00643
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A:Title: C-erbB activation in ALV-induced erythroblastosis: novel RNA processing and processing of the gag-env-erbB fusion
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: B00643
A:Molecule type: mRNA
A:Residues: 1-698 <NIL>
A:Cross-references: GB:M10066; GB:M13881; NID:g211749; PID:AAA48763.1; PID:g211750
A:Note: in Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gallus
C:Comment: This protein is synthesized as a gag-env-erbB protein.
C:Genetics:
A:Gene: gag-env-erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific protein kinase
F:1-6/Product: gag protein (fragment) #status predicted <GAG>
F:7-59/Product: env protein (fragment) #status predicted <ENV>
F:60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F:194-459/Domain: protein kinase homology <KIN>
F:202-210/Region: protein kinase ATP-binding motif
F:229/Active site: Lys #status predicted

Query Match 24.8%; Score 1683.5; DB 1; Length 698;
Best Local Similarity 50.6%; Pred. No. 1.4e-62;
Matches 362; Conservative 82; Mismatches 140; Indels 131; Gaps 20;

QY 578 GPEADQCACAHYKDPFPCVRCFSGVKPFDLSYMPKFPBEGACQPCINCTHSCVDL 637
Db 60 GP--DHCWKCAHFDGPHCVKACGAVLGENDTL-VWKYADANAVCOLCHNCTRGCKGP 116
QY 638 DDKGCAPBQARSPITSVISAVV-GILLVVLGVVFGILII--QYI---KANSKFTGITEL- 690

Db 117 GLEGP--NGSKTPSTAAGVVGGLCLLVVVGIGLYLRRRHIVRKTRRLRLQERELV 173
QY 691 -PLTPSGAMPNQAMRIILKETELKVKVYLGSGAGFTVYKGIWIPDGENVKVIPVAIKVURE 749
Db 174 EPLTPSGEAPNQAHRLILKETEFKKVKVYLGSGAGFTVYKGLWIPGEKVKVIPVAIKELRE 233
QY 750 NTPSKANKEILDEAYVWAGVSPYVSRLLGICLTSTVOLVTQLMYPYCGCLLDHVRNRRGL 809
Db 234 ATSPKANKEILDEAYVWASVDNPHVCRLLGICLTSTVQLITQLMYPYCGCLLDYIREHKONI 293
QY 810 GSQDLNWCQIAKMSYLEDVRLVHRDLAARNVLKSPNHNKVTDFGLARLIDIDETEH 869
Db 294 GSQYLLNWCQIAKGMNVLERRLVHRDLAARNVLVKTTPQHKVITDFGLAKLGADEKEY 353
QY 870 HADGKVPKIKMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIARETIPDLLEK 929
Db 354 HAEGKVPKIKMALESILHRYTHQSDVWSYGVTVWELMTFGSKPYDGIPIASEISSVLEK 413
QY 930 GERLPQPPICITIDVYIMVWCWMTDSECRPRFRELSEFSMARDPQRFVVIQ-NEDLGP 988
Db 414 GERLPQPPICITIDVYIMVWCWMTDADSRPKFRELIAEFSKMARDPPRYLVIOGDERMHL 473
QY 989 ASPLDSTFYRSLLLEDDMDGLVDAEYLVPOQGFCCPDPAFGAGGMVHRRHSSSTRSGG 1048
Db 474 PSPTDSKPYRTLMEEEDMEDIVDAEYLVPHQGF-----NSPST----- 513
QY 1049 GDLTLGLEPSEEAAPRPL-----APSEGAGSDVDFDGLGMAAGLQSLPHTDPSPLQR 1103
Db 514 -----SRPLSSLSATSNNSATNCID-----RNGQGHVPRDSFVQR 551
QY 1104 YSEDPTVPLPSET--DGYVAPLTCSPQPEYVQPDVPRQPPSPREGPLPAARPAAGATLER 1161
Db 552 YSSDPTGNFLEESIDDGFL-----PAPEYVQ--LMPKPS----- 585
QY 1162 AKTLSPGKNGVVKDVF-----AFGGAIVENPEYLTPOGGAAPQPHPPPAFS 1206
Db 586 ---TAMVQNQIYNNISLTAISKLPMSRYQNSHSTAVIDNPEYL-----NTNQSPPLAK 634
QY 1207 PAFDNLVYWDQ-----DPPE-----RGAPPSTFKGTPTAENPEYLGLDVP 1246
Db 635 TVFESSYWIQSGNHQINLNDPVDYQDFLFPNETKPNGLLKVPAEAENPEYLRVAAP 689

RESULT 12
TVYUHV
Protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)
C:Species: avian erythroblastosis virus
C:Date: 18-Apr-1994 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C:Accession: A00644; A38022
R:Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A:Title: The erbB gene of avian erythroblastosis virus is a member of the src gene family
A:Reference number: A00644; MUID:84026539; PMID:6313229
A:Accession: A00644
A:Molecule type: DNA
A:Residues: 1-604 <YAM>
A:Cross-references: GB:X01216; NID:g209676; PID:AAA42400.1; PID:g209678
R:Debuire, B.; Henry, C.; Benaissa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Martin, Science 224, 1456-1459, 1984
A:Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type of protein kinase homology
A:Reference number: A38022; MUID:84223957; PMID:6328658
A:Accession: A38022
A:Molecule type: DNA
A:Residues: 1-28, 'W', 30-139, 'F', 141-145, 'V', 147-152 <DEB>
A:Cross-references: GB:X02006
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific protein kinase
F:130-395/Domain: protein kinase homology <KIN>
F:138-146/Region: protein kinase ATP-binding motif
F:165/Active site: Lys #status predicted

Query Match 23.9%; Score 1620; DB 1; Length 604;
Best Local Similarity 50.5%; Pred.No. 4.9e-60;
Matches 348; Conservative 78; Mismatches 131; Indels 132; Gaps 18;

QY 587 CAHYKPPFCVACPCGVPDLVSMPILWKFPDEEGACOPCINCTHSCVDLDDKGCPAEQ 646
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 3 CAHFIDGPHCVKACAGVLGENTLT-VRKTYADANAVCOLCHPNCRTCKGCGFUEGCP--- 58

QY 647 RASPLTSIVSAVV-GILLVVVLGVVFGLI--QYI---KANSKFIGITEL--PLTSPSGAM 698
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 59 NGSKTFSIAAGVVGGLCLTVVGLGICLYLRHHVRKRTLRLLOERELVELPTPSGEA 118

QY 699 PNOAQWRIKLKETELRKKVLGSGAFGTYYKGIWIPGENVKVIPVAIKVLRENTSPKANKE 758
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 119 PNOAHLRIKLKETEFKKVKVLGSGAFGTYYKGLWIPEGEKVKIPIVAIKELREATSPKANKE 178

QY 759 ILDEAYVMAGVGPYSYRLGICLTSTVOLATOLMYPVGCLLDHHVNRGRFLGSODLLNWC 818
||||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 179 ILDEAYMASVDNPHVCRLGICLTSTVQITQLMYPVGCLLDIHREHKONIGSQYLNNWC 238

QY 819 MQIAGMSYLEDLVRLHARDLAARNLVKS PNHVKITDFGLARLLDDIETEHADGGKVI 878
||||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 239 VQIAGKNVLEERRLVHRDLAARNLVKTPOHVKITDFGLAKLLGADEKEYHAEGKVI 298

QY 879 KWMALSESILRRFTHQSDVMVSYGVTWELMTFGAKPYDGIPAREIPDLLKEGERLPQPI 938
||||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 299 KWMALSESILHRIVTHQSVDVSYGVTWELMTFGSKPYDGIPASEISSVLEKGERLPQPI 358

QY 939 CTIDVYMIMVKCMWDISECPRELVSERSWARDPORFWIQ-NEDLGPA SPLSTFFY 997
||||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 359 CTIDVYMIMVKCMWDIADSPKPRELIAEESKWAPDPRYLVIQDERMELPSPTDSKFY 418

QY 998 RSILEDMDGMLDVDAEYLVPOQFFCPDPAPGAGGMVHRHSRSSSTRSCGGDLTLGLEP 1057
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 419 RTLWEEDMEDIVIDADYLVPHQGFF-----NSPST----- 449

QY 1058 SEEAAPSPL-----APSEGA GDVFDGDZMGAAKGLSLPTHDSPLQRYESDPTVPL 1112
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 450 -----SRTPLLSSLSATSNNSATNCID-----RNQGHVPVREDSFVQRYSSDPTGNF 496

QY 1113 PSET-DGYVAPLTCSPQPVYNQDVRPOPSPREGPLPAARPAGATLERAKTILSPGKN 1170
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 497 LEESIDGDL-----PAPEYVQ--LMPKKPSTAM----- 524

QY 1171 GVVKDVFAP-----GGAIVENPEYLTPOGGAAPQHPPPAFSPAFAFN 1211
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 525 -VNQIYNFISLTAISKLPWDSRYQNSHSTAVDNPEYL-----NTNQSPLA KT VFES 575

QY 1212 LYTWDDQPBERGAPPSTFKGTPTAENPEY 1240
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 576 SPYWIQSGNHQ-----INLDNFDY 594

RESULT 13
GFPE
epidermal growth factor receptor - fruit fly (*Drosophila melanogaster*)
N:Contains: protein-tyrosine kinase [EC 2.7.1.112] erbB
C:Species: *Drosophila melanogaster*
C>Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 11-Jun-1999
C/Accession: A00640; A38021
R:Livneh, B.; Glazer, L.; Segal, D.; Schlessinger, J.; Shilo, B.Z.
Cell 40, 599-607, 1995
A>Title: The *Drosophila* EGF receptor gene homolog: conservation of both hormone binding
A/Reference number: A00640; MUID:85124611; PMID:2982499
A/Accession: A00640
A/Molecule type: DNA
A/Residues: 1-1330 <LIV>
A/Cross-references: EMBL:K03054
R:Wadsworth, S.C.; Vincent III, W.S.; Biodeau-Wentworth, D.
Nature 314, 178-180, 1995
A>Title: A *Drosophila* genomic sequence with homology to human epidermal growth factor re
A/Reference number: A38021; MUID:85137936; PMID:2983232
A/Accession: A38021

A:Molecule type: DNA
A:Residues: 'A',832-866,'V',868-943,'QTPSLVK' <WAD>
A:Cross-references: EMBL:X02293; NID:G7922; PIDN:CAA26157.1; PID:G929565
C:Comment: This sequence is tentative because the introns have not been identified.
C:Genetics:
A:Gene: FlyBase:Egfr
A:Cross-references: FlyBase:FBgn0003731
A:Map position: 2 57F
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F:1-732/Domain: extracellular #status predicted <EXT>
F:733-764/Domain: transmembrane #status predicted <TM>
F:765-1330/Domain: intracellular #status predicted <INT>
F:808-1072/Domain: protein kinase homology <KIN>
F:816-824/Region: protein kinase ATP-binding motif
F:122,300,324,363,518,688,695,700/Binding site: carbohydrate (Aen) (covalent) #status pr
F:774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:843/Active site: Lys #status predicted
F:1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predict

	Query Match	23.8%	Score 1612.5;	DB 1;	Length 1330;
	Best Local Similarity	29.6%;	Pred. No. 2.1e-59;		
	Matches 409;	Conservative 175;	Mismatches 413;	Indels 385;	Gaps 39;
Qy	80	VOGVYLIAHNOVRQVPLQRILVRGTQLF-----EDNYALAVLDNGDPLNNTPTVTGASP	134		
Db	38	ITNYIVIGLDIPTCLSYLQIRGLTIFSLSVEEKALFV-----TY	81		
Qy	135	GLRELQLRSITELKGGVLTORNPOLCVQDTILWKDIHFNKNQLALTLDITNRSRACHP	194		
Db	82	SKMTYLEIPDLRDVLVNGQVGFHNHNLCHMRTIQWSEIVSNGTDAYNYDYFTAPERSCP	141		
Qy	195	CSPMCKSGRCWGESSEDQSLTRTRVCAGGCA--RCKGPLPTDCCHQCQAAGCTGPKHSDC	252		
Db	142	CHESCTHG--CWGEGPKNCKPSKLTCSQCACGRGCVGPARECCHLFCAGCGTGPTQKDC	200		
Qy	253	LACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPNYNLSTDVGSCT	312		
Db	201	IACKNFDEAVSKBCECPMRKYNPTTYVLETPNPGKYAYGATCVKECP--GHLLRDNGACV	259		
Qy	313	LVCPLHNOEVTAEQDQCEKSKPCARVCYGLGMQYIKANSKFIGITEL-----EFAG	366		
Db	260	RSCPQDKMDKGG-----CVPNGCPKTC-----PGVTVLHAGNIDGFRN	300		
Qy	367	CKKTFGLSLAFLPESPDG--DPASNTA-----PLOPEQLQVFETLBEITGYLIVISAWPDS	418		
Db	301	CTVIDGNRIIDQTFSGQDVVANYTMGPRIYPLDPERREVFSTVKEITGYLNEGTHPQ	360		
Qy	419	LPDLSVFQNLQVIRGLIHNGAY--SLTLOGLGSIWGLRSLRELGSGLALIHNNTHLCFV	477		
Db	361	FRNLSPFRNLETIHGRQLMESMAFALAIKSSLSLYSLEMRNLKQISSGVSIVIQRNRLCYV	420		
Qy	478	HTVPWDQLFRNPHQALLHTANRPEDE-----DPPE-----	504		
Db	421	SNIRWPAIQKEPEQKVVYNNENLRADLCGKFLTILISVQHNIIMHIFAICREKNHLLGSV	480		
Qy	505	-----	504		
Db	481	QRGLLGSWHGVSPLYQLQLQFQWHLHRRLLWLYIQVINSITQDSNEHQLTDACYSPSVPT	540		
Qy	505	-----	506		
Db	541	SLTIERARYAQTOSAGLAMELEQITARSAMRHSKTLPAEGRQVPRWVFLGVCASARAGIA	600		
Qy	507	EGLA-----CHQICARGHCWGPQPTQVNCISOFLRGECVEECRVLOGLPREYV---N	556		
Db	601	EPLAGRAVCRKCHPLCELCTNNGYHEQVCSKCTHYKRREQCETEC-----PADHYTDBE	654		
Qy	557	ARHCLPCHPEEQPQNGSVTCFGEPAQDCVACAHYK-----DPPE-----CVARCPSG	603		
Db	655	QRECFQRHPEC---NG---CTGFGADDCKSCRNFKLFANETGTYVNSTWNCFTSKCPLE	708		
Qy	604	VK--PDLSYMPIWKPFDEEGACQPCINCTHSCVDLDDKGCPCAPQRASPLTISVSAVVGIL	662		

```
Db 709 MRHVNQTAIGPY-----CAASPPRSSKITANLDVN-----MFIITGAVLVP 752
Qy 663 LVVLGVVFGILIOYIKANSFIGIT-----ELPLTPSGAMPNQAMRILKETELRK 714
Db 753 TICILCVVYICROKQAKKETVMTALSGREDEPLRPSNIGANLCKLRIVKDAELRK 812
Qy 715 VKVLGSGAGFTYKGIWIIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVAGVGSYV 774
Db 813 GGVLGMGAFGRYKGVWVPEGENKIPVAIKELLKSTGAESSEEPLEAYIMASEEHWNL 872
Qy 775 SRLIGICLTSTVOLTPQMLPBYCGLLDHVNRGLRGSODLLNWCQIAKGSYLEDVPLV 834
Db 873 LKLLAVCMSSQMLITQMLPGCLLDYVYRNRDKTIGSKALLNWSQIAKGSYLEEKRLV 932
Qy 835 HRDLAARNVLVK---SPNHVITDIFGLARLLDIDETEHADGGKVPKIKWMALESTLRRF 891
Db 933 HRDLAARNVLRLLAGEDH---DFGLAKLLSSDSNEYKAAGKWPKIKWLALECIERNVF 988
Qy 892 THQSDVWSYGVTVWELMTFGAKPYDGPAPRIPDLLEKGERLPQPPICTIDVYIMVKW 951
Db 989 TSKSDWAFGVTIWELLTFGORPHENIPAKDIPDLIEVGLKLEQPEICSLDIYCTLLSCW 1048
Qy 952 MIDSECRPRFRELVSFSESMARDPQRFVVIQNEDLG--PASPLOSTFVRSILLEDD---DM 1006
Db 1049 HLDAAARFTFKOLTTVFAEFARDPGRYLAILGDKFTRLPA-----YTSODEKDLIRKL 1101
Qy 1007 GDLVDAEYLVPOQGFPCDPAPGAGGMVHRHSSSTRSGGDLTLGLEPSEEP--- 1063
Db 1102 APTTDGSAIAKAPDDYLOPKAALGPS-----HRTDCT-----DEMPKLN 1140
Qy 1064 ---RSLAPSGAGSDVFDG---DLGMAAGKLOSLPHTDPSPLQVSEDPVPLPSETD 1117
Db 1141 RYCKDPSKNKSTGDDERDSSAREVGVGNLR-----LDLPVDED 1179
Qy 1118 GYVAPLTCSPQEVYNQDVRPQPPSPREGPLPAARPAATILERAKTLSPOKNGVVKDVF 1177
Db 1180 DYLMPT-TCQPGNNNNNNV-----NPNQNNAAVGVAAAGYM-----DLI 1217
Qy 1178 AFGGAVENPEYL---TPQGGAAPOPH-----PPAPSP-APFDNLYY 1214
Db 1218 GVPVSVNDPEYLLNAQTILGVSEPIPTQITIGIPVWGPGTMEVKVMPGSEPTSSDHEY 1277
Qy 1215 WD 1216
Db 1278 ND 1279

RESULT 14
S35745
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
C:Species: avian erythroblastosis virus
C:Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
C:Accession: S35745
R:Vennstrom, B.
submitted to the EMBL Data Library, March 1993
A:Reference number: S35743
A:Accession: S35745
A:Molecule type: DNA
A:Residues: 1-544 <VEN>
A:Cross-references: EMBL:X12707
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP, oncogene, phosphotransferase; transforming protein; tyrosine-specific P
F:135-400/Domain: protein kinase homology <KIN>
F:170/Active site: Lys #status predicted

Query Match 23.1%; Score 1564; DB 2; Length 544;
Best Local Similarity 53.1%, Pred. No. 9e-58;
Matches 333; Conservative 72; Mismatches 124; Indels 98; Gaps 17;
```

```
Qy 578 GPEADQCVACAHYKDPFPFCVARCPSGVKPDLISYMPIWKFPDEBGAQCPCPINCTHSCVDL 637
Db 1 GP--DHCWKCAHFIDGPHCVKACAPAGVLGENDTL--VMKYADANAVCOLCHPNCRTGCKGP 57
Qy 638 DDGCPRAEQRASPLTSTVSAVV--GILLVVLGVVFGILI--QYI---KANSKFGITEL- 690
Db 58 GLEGP---NGSKTPSTAAGVVGGLCLLVVVGIGLYLRRRHIVRKRITLRLRLLQERELV 114
Qy 691 -PLTPSGAMPNQAMRILKETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRE 749
Db 115 EPLTPSGEAPNQAHRLILKETEFKVKVLGFGAGFTYKGLWPEGEKVTIPVAIKELRE 174
Qy 750 NTSPKANKEILDEAYVAGVGSYPVSRLLGICLTSTVOLTPQMLPBYCGLLDHVNRGRRL 809
Db 175 ATSPKANKEILDEAYVNASVDNPHVCRLLGICLTSTVOLITQMLPBYCGLLDYIREHKDNI 234
Qy 810 GSODLLNWCQIAKGSYLEDVRLVHRDLAARNVLKSPNHVKITDIFGLARLLDIDETEH 869
Db 235 GSQYLLNWCQIAKGMNYSLEERHMHVRLDLAARNVLVKTPOHKITDIFGLAKOLGADEKEY 294
Qy 870 HADGGKVPKIKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGPAPRIPDLLEK 929
Db 295 HADGGKVPKIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGPAPSEISSVLEK 354
Qy 930 GERLPQPPICTIDVYIMVKWMIDSECRPRFRELVSFSESMARDPQRFVVIQ-NEDLGP 988
Db 355 GERLPQPPICTIDVYIMVKWMSDADSRPKFRELIAEFSSKWARDPPRYLVIQDERMHL 414
Qy 989 ASPLDSTFYRSILLEDDMGDLVDAEYLVPOQGFPCDPAPGAGGMVHRHSSSTRSGG 1048
Db 415 PSPDTSKPYRTLMBEEDMEDIVDAEYLVPHQGF-----NSPST----- 454
Qy 1049 GDLTLGLEPSEEPSEAPRSP-----APSEGAGSDVFDGDLGMAAGKLOSLPHTDPSPLQR 1103
Db 455 -----SRPFLSSLSATSNSNATCIDRNGG-----H----- 481
Qy 1104 YSEDPVPLPSETDGYVAPLTCSPQEVYNQDVRPQPPSPREGPLPAARPAATILERA 1162
Db 482 -----PYREDGFL-----PAPEYVQ--LMPKKPSTAMVQNIYNYISLTAISKL 524
Qy 1163 KTLSPGKNGVVKVAFPGGAVENPEYL 1189
Db 525 PIDSRYQN-----SHSTAVDNPEYL 544

RESULT 15
S00727
kinase-related transforming protein (erbB) (EC 2.7.1.1-) - avian erythroblastosis virus
C:Species: avian erythroblastosis virus
C:Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
C:Accession: S00727
R:Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
Oncogene Res. 1, 265-278, 1987
A:Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mutant
A:Reference number: S00727; MUID:88217326; PMID:2897102
A:Accession: S00727
A:Molecule type: DNA
A:Residues: 1-545 <SCO>
A:Cross-references: EMBL:X06943
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; phosphotransferase
F:135-400/Domain: protein kinase homology <KIN>
F:143-151/Region: protein kinase ATP-binding motif

Query Match 23.0%; Score 1557; DB 2; Length 545;
Best Local Similarity 53.1%, Pred. No. 1.8e-57;
Matches 333; Conservative 71; Mismatches 125; Indels 98; Gaps 17;
```

```
Qy 578 GPEADQCVACAHYKDPFPFCVARCPSGVKPDLISYMPIWKFPDEBGAQCPCPINCTHSCVDL 637
Db 1 GP--DHCWKCAHFIDGPHCVKACAPAGVLGENDTL--VMKYADANAVCOLCHPNCRTGCKGP 57
```

```
QY 638 DDKCPAQRASPLTSIYSAVV-GILLVVVLGVVFGILI--QYI---KANSKFIGITEL- 690
Db 58 GLEGCP---NGSKTPIAAGVVGGLCLVVLGGLYGLYRRRHIVRKTLRLRLOERELV 114
QY 691 -PLTPSGAMPNOAOMRIKTELKRVKVLGSGAFCTVYKGIWIPDGENVKIPVAIKVIRE 749
Db 115 EPLTPSGEAPNOAHLRIKTEFEKKVKVLGFGAFCTVYKGLWIPGEKVTIPVAIKELRE 174
QY 750 NTSPKANKEIIDEAYVMAVGSPYVSRLGICLTSTVQLVTLMPYGCGLDHHVRENRL 809
Db 175 ATSPKANKEIIDEAYVMAVDNPHVCRLLGICLTSTVQLITQLMEPYGCLLDYIREHKDNI 234
QY 810 GSODLLNMCMOIAKMSVLEDRVLVHRDLAARNVLAKSPNHVKITDFGLARLLDIDETEY 869
Db 235 GSQYLLNMCVQIAKMNYLEERHLVHRDLAARNVLVKTPODVKITDFGLAKQLGADKEY 294
QY 870 HADGGKVPKWMALESIILRRFTHQSDVMSYGVTVWELMTFCAPYDGIPIAREIPDLLEK 929
Db 295 HAEKKVPKWMALESIILHRYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISSVLEK 354
QY 930 GERLPQPPICTIDVTMIMVKMIDSECRPRPRELVSEFSRMARDPQRFVVIQ-NEDLGP 988
Db 355 GERLPQPPICTIDVTMIMVKMWDADSRPKFRELIAEFKWARDPPRYLVITQGERMHL 414
QY 989 ASPLDSTFYRSILLEDDMDGLVDAREYLVPOGGFPCDPAPGAGGVHHRHRSSTRSGG 1048
Db 415 PSPTDSKFYRTLMEEDMEDIVDADEYLVPHQGF-----NSPST----- 454
QY 1049 GDLTLGLEPSEBEAPRSP-----APSEAGSDVDFDGLGMAKGLQSLPHTDPSPLQR 1103
Db 455 -----SRTPLLSLSATSNNATNCIDRNG-----H----- 481
QY 1104 YSEDTVPLPSETDGYVAPLTCSPQPEYVNPQDVPRPQPPSPREGPLPAARPAGAT-LERA 1162
Db 482 -----PVREDGFL-----PAPEYVNO--LMPKKPSTAMVQNIYVISTAIKSL 524
QY 1163 KTLSPKNGVVKDVFAGGAVENPEYL 1189
Db 525 PMDSRYQN-----SHSTAVDNPEYL 544
```

Search completed: July 22, 2003, 09:27:34
Job time : 28.1435 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:10:23 ; Search time 22.2855 Seconds
(without alignments)
5413.772 Million cell updates/sec

Title: SEQ4-695-709-12
Perfect score: 6815
Sequence: 1 MEALACRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6640	97.4	1255	1 A24571	protein-tyrosine k
2	5843	85.7	1260	1 TVRTWU	protein-tyrosine k
3	5833.5	85.6	1254	2 I48161	p-185 precursor -
4	3067	45.0	1210	1 GQHUE	epidermal growth f
5	3038	44.6	1210	2 A53183	epidermal growth f
6	3014.5	44.2	1223	1 TVCHLV	epidermal growth f
7	2887.5	42.4	1308	2 A47253	epidermal growth f
8	2598	38.1	1166	1 S06142	protein-tyrosine k
9	2332.5	34.2	1342	2 A36223	kinase-related tra
10	2246.5	33.0	1339	2 JC4387	epidermal growth f
11	1686.5	24.7	698	1 TVFVLV	protein-tyrosine k
12	1623	23.8	604	1 TVYUHV	protein-tyrosine k
13	1575	23.1	544	2 S35745	protein-tyrosine k
14	1568	23.0	545	2 S00727	kinase-related tra
15	1565.5	23.0	1330	1 GQFFE	epidermal growth f
16	1551	22.8	540	2 B44776	protein-tyrosine k
17	1549	22.7	540	1 TVFVVB	protein-tyrosine k
18	1509	22.1	644	2 A36325	epidermal growth f
19	1243	18.2	1323	2 E88257	protein let-23 (im
20	1243	17.0	1374	2 S70712	protein-tyrosine k
21	1160	17.0	1369	2 S70713	protein-tyrosine k
22	1134	16.6	1717	1 A45558	epidermal growth f
23	1126	16.5	527	2 A42032	epidermal growth f
24	975.5	14.3	843	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	protein-tyrosine k
26	754.5	11.1	311	2 S13808	protein-tyrosine k
27	700	10.3	1363	2 T43220	insulin-like growth
28	666	9.8	1383	2 A36080	insulin receptor p
29	665.5	9.8	1372	2 A34157	insulin receptor p

RESULT 1

A24571

Protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N/Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein erbB

C/Species: Homo sapiens (man)

C/Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999

C/Accession: A24571; A25491; A44188; B44188; I59509; I57622

R/Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;

Nature 319, 230-234, 1986

A/Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth f

A/Reference number: A24571; MUID:86118663; PMID:3003577

A/Accession: A24571

A/Molecule type: mRNA

A/Residues: 1-1255 <YAM>

A/Cross-references: GB:X03363; NID:G31197; PIDN:CAA27060.1; PID:G31198

R/Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.

Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A/Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epiderm

A/Reference number: A25491; MUID:86016729; PMID:2995967

A/Accession: A25491

A/Molecule type: DNA

A/Residues: 737-1031 <SEM>

A/Cross-references: GB:M11767; NID:G182163; PIDN:AAA35808.1; PID:G53282

R/Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg, P.

Science 230, 1132-1139, 1985

A/Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromos

A/Reference number: A44188; MUID:86070181; PMID:2999974

A/Accession: A44188

A/Molecule type: DNA

A/Residues: 740-910 <COU1>

A/Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989

A/Accession: B44188

A/Molecule type: mRNA

A/Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>

A/Cross-references: GB:M11730; NID:G183986

R/King, C.R.; Kraus, M.H.; Aaronsen, S.A.

Science 229, 974-976, 1985

A/Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A/Reference number: I59509; MUID:85272597; PMID:2992089

A/Accession: I59509

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 832-909 <REX>

A/Cross-references: GB:L29395; NID:G459807; PIDN:AAA35809.1; PID:G459808

R/Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987

A/Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptional

A/Reference number: I57622; MUID:87286898; PMID:3039351

A/Accession: I57622

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-191 <TAL>

A;Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C;Genetics:
A;Gene: GDB:ERBB2; NGL; NEU; HER-2
A;Cross-references: GDB:120613; OMIM:164870
A;Map position: 17q21.1-17q21.1
A;Introns: 25/1; 75/3; 147/1; 883/3
A;Note: the list of introns is incomplete
C;Function:
A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
inase
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F;22-653/Domain: extracellular #status predicted <EXT>
F;70-304/Domain: EGF receptor extracellular domain repeat <EE1>
F;395-605/Domain: EGF receptor extracellular domain repeat <EE2>
F;654-675/Domain: transmembrane #status predicted <TM>
F;676-1255/Domain: intracellular #status predicted <INT>
F;718-983/Domain: protein kinase homology <KIN>
F;726-734/Region: protein kinase ATP-binding motif
F;168,124,187,259,530,571,629/Binding site: carboxylate (Asn) (covalent) #status predicted
F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;753/Active site: Lys #status predicted
F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 97.4%; Score 6640; DB 1; Length 1255;
Best Local Similarity 97.5%; Pred. No. 5.5e-266;
Matches 1255; Conservative 8; Mismatches 21; Indels 2; Gaps 2;

QY 1 MELAAALCRWGLLLALLPPGAASCTGCTDMKRLRASPETHDMLRLHYOGCQVQGNL 60
DB 1 MELAAALCRWGLLLALLPPGAASCTGCTDMKRLRASPETHDMLRLHYOGCQVQGNL 60
QY 61 ELTYLPTNASLSFLQDIEVOGYVLIHNAQVRQVPLQRLRVGRQLFEDNYALAVLNG 120
DB 61 ELTYLPTNASLSFLQDIEVOGYVLIHNAQVRQVPLQRLRVGRQLFEDNYALAVLNG 120
QY 121 DPLNTTPTVTGASPGGLRELQRLSLEILKGGVLIQRNPOLCYQDTILWKDI FHKNNOLA 180
DB 121 DPLNTTPTVTGASPGGLRELQRLSLEILKGGVLIQRNPOLCYQDTILWKDI FHKNNOLA 180
QY 181 LTLIDTNRSRACHPCSPCKSGRWGESSEDCQSITRTVCAGGCARCKGLPTDCCHQC 240
DB 181 LTLIDTNRSRACHPCSPCKSGRWGESSEDCQSITRTVCAGGCARCKGLPTDCCHQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAC 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAC 300
QY 301 YNLTSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLGMQYIKANSKFTGIT 360
DB 301 YNLTSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLGMQYIKANSKFTGIT 360
QY 361 ELEFAGCKKIFGSLAFLPESPDGDPASNTAPLOPELOVFEITLGYLVIASWPDLSL 420
DB 361 IQEFAGCKKIFGSLAFLPESPDGDPASNTAPLOPELOVFEITLGYLVIASWPDLSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLSRLRELGSGLALIHNNTHLCFVHTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLSRLRELGSGLALIHNNTHLCFVHTV 480
QY 481 PWDQLFRPHQALLTANRPEDECVGEGLACHOLCARGHCWGPPTQVCNCSQFLRGQEC 540
DB 481 PWDQLFRPHQALLTANRPEDECVGEGLACHOLCARGHCWGPPTQVCNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREYVNAHRLCPCHPECOPOGNSVTCFGEADQCACAHYKDPPEVCARC 600
DB 541 VEECRVLQGLPREYVNAHRLCPCHPECOPOGNSVTCFGEADQCACAHYKDPPEVCARC 600
QY 601 PSGVKPDLISYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTISAVVG 660
DB 601 PSGVKPDLISYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTISAVVG 660

601 PSGVKPDLISYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTISAVVG 660
QY 661 ILLVVLGVVFGIILIKRRQKIRKYTWRRLLQETVELPELTPSGAMPNQAMRILKEQYI 720
DB 661 ILLVVLGVVFGIILIKRRQKIRKYTWRRLLQETVELPELTPSGAMPNQAMRILKEQYI 720
QY 721 KANSKFITITEL-TVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVGS 779
DB 721 R-KVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVGS 779
QY 780 PYVSRLLGICLTSTVQLVTQMLPVGCLLDHVNRNRLGSDLLNWCQAKGMSYLEDV 839
DB 780 PYVSRLLGICLTSTVQLVTQMLPVGCLLDHVNRNRLGSDLLNWCQAKGMSYLEDV 839
QY 840 RLVRHDLAARNVLKSPNHVKITDFGLARLLDIDETVHADGGKVPKKNWALESLILRRRF 899
DB 840 RLVRHDLAARNVLKSPNHVKITDFGLARLLDIDETVHADGGKVPKKNWALESLILRRRF 899
QY 900 THQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMVMKWC 959
DB 900 THQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMVMKWC 959
QY 960 MIDSECRPRELVSEFSRWARDPQRVVIQNEIDLGSPASPLDSTFYRSLLDEDDMGDLVD 1019
DB 960 MIDSECRPRELVSEFSRWARDPQRVVIQNEIDLGSPASPLDSTFYRSLLDEDDMGDLVD 1019
QY 1020 AEEVLVPOQGFCCPDPAAGGVMVHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSE 1079
DB 1020 AEEVLVPOQGFCCPDPAAGGVMVHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSE 1079
QY 1080 GAGSDVFDGDLGGAAGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEY 1139
DB 1080 GAGSDVFDGDLGGAAGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEY 1139
QY 1140 VNQDVRPQPPSPREGPLPAARPAATLERAKTLSPQKNGVYKDVFAFGGAVENPEYLTP 1199
DB 1140 VNQDVRPQPPSPREGPLPAARPAATLERAKTLSPQKNGVYKDVFAFGGAVENPEYLTP 1199
QY 1200 QGGAAPHPHPPAPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
DB 1200 QGGAAPHPHPPAPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 2
TVRTNU
Protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
C;Accession: A24562; A61204
R;Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A;Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.
A;Reference number: A24562; MUID:86118662; PMID:3945311
A;Accession: A24562
A;Molecule type: mRNA
A;Residues: 1-1260 <BAR>
A;Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746
R;Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen, C.
Carcinogenesis 12, 1975-1978, 1991
A;Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no m-2-thiazolyl]formamide or N-methyl-N-nitrosourea.
A;Reference number: A61204; MUID:92035293; PMID:1682063
A;Accession: A61204
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 637-663, 'V', 665-702 <MAS>
A;Note: authors translated the codon GCA for residue 25 as Val
C;Genetics:
A;Gene: neu
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

QY 361 ELEFAGCKKIFGSLAFLESFDPASNTAPLOEQLQVFTFLBEITGLYLSAWPDSLP 420
DB 361 IOEFAGCKKIFGSLAFLESFDPGPNSSGIAFLTEQLQVFTFLBEITGLYLSAWPDSLH 420
QY 421 DLSVFQNLQVIRGRIHNGAYSLSLQGLIGISWGLRLSRLRELSGLALHFNTHLFCFVHTV 480
DB 421 DLSVFQNLQVIRGRIHNGAYSLSLQGLIGISWGLRLSRLRELSGLALHFNTHLFCFVHTV 480
QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPGPTQCVNCSQFLRGQEC 540
DB 481 PWDQLFRNPHQALLHSGNFSSEECGLKDFACYPCLCAHGHCHWGPGPTQCVNCSHFLRGQEC 540
QY 541 VEECRVLQGLPREYVNAHCHLPCHEPCOPNGSVTCFGEADQCVACAHYKDPFCFVARC 600
DB 541 VKCRVWKLPREYVNGHCHLPCHEPCOPNGSVTCFGEADQCVACAHYKDPFCFVARC 600
QY 601 PSQVKPDLSPYPIWKFPPDEEGACOPCINPCTHSCVDLDKGCAPAEQASPLTSIVSAVVG 660
DB 601 PSQVKPDLSPYPIWKYPDEEGACOPCINPCTHSCVDLDKGCAPAEQASPLTSIATVVG 660
QY 661 ILLVVVLGVFGLIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAOMRILKEQYI 720
DB 661 ILLFLVIGVVGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAOMRILKETEL 720
QY 721 KANSKFIGITEL-TVYKGIWIPDGENVKIPVAIKVLRNTPSKANKETLDEAYVMAGVGS 779
DB 721 R-KVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKETLDEAYVMAGVGS 779
QY 780 PYVSRLLGICLTSTVOLQTMPLYGCLLDHVRENRGLGSDLLNWCQIAKGSYLEDV 839
DB 780 PYVSRLLGICLTSTVOLQTMPLYGCLLDHVRENRGLGSDLLNWCQIAKGSYLEDV 839
QY 840 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRRF 899
DB 840 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRRF 899
QY 900 THQSDVMSYGVTVWELMTFGAKPDYDGPAREIPDLLEKGERLPPOPICTIDVYIMWKWCW 959
DB 900 THQSDVMSYGVTVWELMTFGAKPDYDGPAREIPDLLEKGERLPPOPICTIDVYIMWKWCW 959
QY 960 MIDSECRPRFRELSEFSEMRARDPQFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDVLD 1019
DB 960 MIDSECRPRFRELSEFSEMRARDPQFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDVLD 1019
QY 1020 AEEVLVPOQGFCDPAPGAGMWHRRSSSTRSGGGDLTLGLPESEEAAPRSPPLAPSE 1079
DB 1020 AEEVLVPOQGFCDPAPGAGMWHRRSSSTRSGGGDLTLGLPESEEAAPRSPPLAPSE 1079
QY 1080 GAGSDVFDGLGMGAAGKQSLPTHPDPSLQRYSEDPTVPLPSETDGVVAPLTCSPQPEY 1139
DB 1080 GAGSDVFDGLGMGAAGKQSLPRLDPSLQRYSEDPTVPLPSETDGVVAPLTCSPQPEY 1139
QY 1140 VNQDVRPQPPSPREGPLFAARPGATLRLAKTLPSPKNGVGVKQVAFGGAVENPEYLTP 1199
DB 1140 VNQDVRPQPPSPREGPLFAARPGATLRLAKTLPSPKNGVGVKQVAFGGAVENPEYLTP 1199
QY 1200 QGGAAPOHPHPPAPSPAFDNLVYWDQDPPERGAPPSTFKGPTTAENPYLGLDVPV 1255
DB 1200 RGGASQPH-PPALCPAFDNLVYWDQDPPSERGSPNTFEGTPTAENPYLGLDVPV 1254

RESULT 4
COHUE
epidermal growth factor receptor precursor - human
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C:Species: Homo sapiens (man)
C:Date: 15-Nov-1994 #sequence revision 27-Nov-1995 #text change 11-Jun-1999
C:Accession: A00641; A25772; S30024; A38672; A06642; A43615; A23062; A05281; A60143; A3319; J. Ullrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.; Yrg, P.H.
Nature 309, 418-425, 1984
A:Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression of

A:Reference number: A00641; MUID:84219729; PMID:6328312
A:Accession: A00641
A:Molecule type: mRNA
A:Residues: 1-1210 <ULL>
A:Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA5240.1; PID:g553272
A:Note: the authors translated the codon AAG for residue 540 as Asn
R:Ishii, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
A:Title: Characterization and sequence of the promoter region of the human epidermal growth factor receptor gene
A:Reference number: A25772; MUID:85270438; PMID:2991899
A:Accession: A25772
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-29 <ISH>
A:Cross-references: GB:M11234; NID:g181981; PIDN:AAA52370.1; PID:g553272
R:Haley, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ullrich, A.; Waterfield, M.
Oncogene Res. 1, 375-396, 1987
A:Title: The human EGF receptor gene: structure of the 110 kb locus and identification of the promoter
A:Reference number: S30024; MUID:88217333; PMID:3329716
A:Accession: S30024
A:Molecule type: DNA
A:Residues: 1-29 <HA2>
A:Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
R:Haley, J.D.; Waterfield, M.D.
J. Biol. Chem. 266, 1746-1753, 1991
A:Title: Contributory effects of de Novo transcription and premature transcript termination on the expression of the human epidermal growth factor receptor cDNA
A:Reference number: A38672; MUID:91107677; PMID:1988448
A:Accession: A38672
A:Molecule type: DNA
A:Residues: 1-29 <HAL>
A:Cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271
A:Experimental source: A431 human carcinoma cell line A431-7
R:Xu, Y.; Ishii, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.; Merlino, G.T.; Pastan, I.
Nature 309, 806-810, 1984
A:Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RNAs
A:Reference number: A00642; MUID:84245835; PMID:6330563
A:Accession: A00642
A:Molecule type: mRNA
A:Residues: 150-187, 'KSIVQAV', 195, 'M', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-321
A:Experimental source: epidermoid carcinoma cell line A431
R:Simmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.
Biochem. Biophys. Res. Commun. 124, 125-132, 1984
A:Reference number: A23062; MUID:85046483; PMID:6093780
A:Accession: A23062
A:Molecule type: mRNA
A:Residues: 713-964 <LIN>
A:Experimental source: epidermoid carcinoma cell line A431
R:Simmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.
Biochem. Biophys. Res. Commun. 124, 125-132, 1984
A:Reference number: A23062; MUID:85046483; PMID:6093780
A:Accession: A23062
A:Molecule type: mRNA
A:Residues: 1028-1210 <SIM>
R:Weber, W.; Gill, G.N.; Speiss, J.
Science 224, 294-297, 1984
A:Reference number: A05281; MUID:84172183; PMID:6324343
A:Accession: A05281
A:Molecule type: protein
A:Residues: 25-30, 'S', 32-51, 454-467 <WEB>
R:Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.
J. Biol. Chem. 260, 5205-5208, 1985
A:Title: Identification of residues in the nucleotide binding site of the epidermal growth factor receptor
A:Reference number: A60143; MUID:85182650; PMID:2985580
A:Accession: A60143
A:Molecule type: protein
A:Residues: 740-744, 'X', 746-747 <RUS>
R:Mroczkowski, B.; Mosig, G.; Cohen, S.
Nature 309, 270-273, 1984
A:Title: ATP-stimulated interaction between epidermal growth factor receptor and supercoiled DNA
A:Reference number: A38023; MUID:84191554; PMID:6325948
A:Contents: annotation; receptor activity

Db 123 YQK-NPSSP--DVTQVGLKQLQLNLNLTEILSGGVKVSHPNPLLCNVETINWWWDIVDKTSNP 179

Qy 180 ALTLIDNRSRACHPCSMCKSGRCWGESSEDQSLTRTVCAAGC-ARCKGPLETDCCHE 238

Db 180 TMNLI PHAFERQCKCDHGCVNGCWAPGPGHCQKFTKLLCAEQCNRRRCRGPKEIDCNE 239

Qy 239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA 298

Db 240 HCAGGCTGPRATDCLACRDFNDGCTCKDTPPPKIYDIVSHVQVVDNPNIKYTFGAACVKE 299

Qy 299 CPYNVLSTDVGSCTLVCPHLHNOEYTAEDGTORCEKCKPCARVCYGLGMQVIKANSKF 358

Db 300 CPSNVVTE-GACVRSACSAGMLEVD-ENGRKSCPCDGVCKVCDGIGISL-SNTIAVN 356

Qy 359 ITEL-EFAGCKKIFGSLAFLESPDGPASNTAPLOPEQLQVFTLEETIGYLVISAWPD 417

Db 357 STNIRSFNCTKINGDIILNRNSPEGDPHYKIGTMDPEHLWNLTTVKEITGYLVIMWPE 416

Qy 418 SLPLSVFQNLQVTRGRILHNGAYS-LTQGLGISWLGRLSLRELGLSLALIHNTLHCF 476

Db 417 NMTSLSVFQNLQVTRGRITFRSGFSFVVVQVRHLQWLGLRLSKEVSAGNVILKNTLQRY 476

Qy 477 VHTVWDQLRPNHQALLHTANRPEDECVGSLACHQL-CARGHCWGPCTOCVNCQFLR 536

Db 477 ANTINWRFLRSEDSQSEYDART-----ENQTCNNECEDSGCW-PEPTMVCVSLHVD 528

Qy 537 GQECVEECRVLQGLPREYNARHCLPCHPECPQNGSVTCFPGPADOCVCAHAKYDPPFC 596

Db 529 GGRCVASCNLLQGPREAQVDGRCVQCHQECVLQVTDLSLTCVPGPANCKSAHFQDGPQC 588

Qy 597 VARPSPGVKPDLSYMPYKWFDPDEGACQPCPINCTHSCVDLDDKCPAERQASPLTIVS 656

Db 589 IPRCPHGILGDGTL-IWKYADKMGCQCPCHQNCCTQCGSGPLSGCRGD-IVSHSLAVG 646

Qy 657 AVVGLLVVVLGVVFGILIKERQKIRYKTRRLLQETELVEPLTPSGAMPNQAMRLK 716

Db 647 LVSGLLTIVALLIVLLRRRRIRK-RKRTIRCLLQELVELPEPTPGQAPNQAFRLIK 705

Qy 717 EOYIKANSKFTIGTEL-TVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMA 775

Db 706 ETEFK-KDRVLGSGAFGTVYKGLWNPDGENIRIPVAIKVLEATSPKVNQEVLEAYVMA 764

Qy 776 VGSPFYVSRLLIGICLTSTVQLVTQMPYGCGLLDHVRENRLGSGQDLLNWCWQIAKMSY 835

Db 765 SVDHPHYCRLLIGICLTSAVQLVTQMPYGCGLLDVVRQHOERICQWLLNWCVQIAKMNY 824

Qy 836 LEDVRLVHRDLAARNVLKSPNHVKITDPGLARLLDDETHYADGGKVPKKNWALSIL 895

Db 825 LEERHLVHRDLAARNVLKSPNHVKITDPGLARLLDDETHYADGGKVPKKNWALSIL 884

Qy 896 RRRPTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGBRLPOPPICTIDVYIM 955

Db 885 QWYTHOSDVMSYGVTVWELMTFGSKPYDGIIPAREIASVLENGERLPOPPICTIEVYMI 944

Qy 956 VKCWMIDSECRPRRELVSFSEARMARPPQFVVIQNEEDLGPASPLDSTFYVRSLLDDMG 1015

Db 945 LKCNMIDPSSRPRRELVSFSEARMARPPQFVVIQNEEDLGPASPLDSTFYVRSLLDDMG 999

Qy 1016 DLVDAEEVLVPOQFFCPDPAFGAGGVHHRHSSSTFSGGGDITLGLFSEEEAPRSL 1075

Db 1000 DVIDADEYLLPYKRI-----NRQGS-----EPC 1022

Qy 1076 APSEGAGSDVPDGLGMAAGLQSLPHTDPSLQRYSEDPV-PLPSETDGVVAPLTC 1134

Db 1023 IPPTGH-----PVRENSITLRNISDPTQNALEKLDGH----- 1055

Qy 1135 POPEYVNPQDVRPQ-----PSPRE-----GPLP-AARPAGATLERAKTLPFGKNGV 1181

Db 1056 ---EYVNPQGSSTSRLLSDIYNPNVEDLTDGWPVPSLSQSAETNFSRPEYLVNTNQSL- 1111

Qy 1182 KDVFAGGAVENPEYLPQGGAAFPQHPAFSPAFNLYWDOQDPPERGAAPSTFKGTP 1241

Db 1112 --PLVSSGMDPPDY---QAG-----YQAAF-----LPQTGALTGNMFLP 1147

Qy 1242 TAENPEYL 1250

Db 1148 AAENLEYIG 1156

RESULT 9

A36223

C:Species: Homo sapiens (man)

C>Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000

C/Accession: A36223; I59164

R/Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.

Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989

A>Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal g

A/Reference number: A36223; MUID:90083234; PMID:2687875

A/Accession: A36223

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-1342 <KRA>

R/Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, G.J

Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990

A>Title: Molecular cloning and expression of another epidermal growth factor receptor-rel

A/Reference number: I59164; MUID:90311312; PMID:2164210

A/Accession: I59164

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>

A/Cross-references: GB:M34309; NID:G183990; PIDN:AAA35979.1; PID:G306841

C/Genetics:

A/Gene: GDB:ERBB3; HER3

A/Cross-references: GDB:119880; OMIM:190151

A/Map position: 12q13-12q13

C/Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog

C/Keywords: ATP; phosphotransferase

F/707-972/Domain: protein kinase homology <KIN>

F/715-723/Region: protein kinase ATP-binding motif

Query Match 34.2%; Score 2332.5; DB 2; Length 1342;

Best Local Similarity 39.5%; Pred. No. 7.5e-99;

Matches 520; Conservative 197; Mismatches 461; Indels 137; Gaps 36;

Qy 10 GLLALLPPGAA--STQVCTGTDMLKRLPASPEETHLMLRHLHYQGCQVQGNLELYLPT 67

Db 11 GLLFSLARGSEVGNQAVCPGTLNGLSVTGAENQYQTLKYERCEVVMGNLEIVLTGH 70

Qy 68 NASLSFLQDIQEVQGVVLIHNOVRQVPLQRLRIVRGTLQFEDNYALAVLDNGDPLNNT 127

Db 71 NADLSFLQWIKFEVTGYVLVAMNEFSTLPLNLRVVRGTQVYDGKFAIVM-----LYNT 125

Qy 128 PVTGASPGGLRELQARSLTEILKGGVLIQORNPOLCYODTILWKDIFHKNNQALTLIDTN 187

Db 126 ----NSSHALRQURLTQLTTEILSGGVYIEKNDKLCMDITDWRD---AEIVVKD 178

Qy 188 RSRACHPCSPMCKSGRCWGESSEDQSLTRTVCAAGC-ARCKGPLETDCCHECAAGCTG 246

Db 179 NGRSCPPCHEVCKG-RCWGPSEDCQTLTKTICAPQCNHGFPGPNQCCHECAGCGSG 237

Qy 247 PKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLST 306

Db 238 PQDITDCFACRHFNDGACVPRCPQPLVYNKLTQLEPNPHTKYQYGGVCVASCAPHNFV-V 296

Qy 307 DVGSCSTLVCPHLHNOEYTAEDGTORCEKCKPCARVCYGLGMQVIKANSKP--IGITELE- 363

Db 297 DQTSQVRCAPPDQKMEVD-KNGLKMCPCPGGLCPKACEGTG-----SGSRFQTVDSNNIDG 350

Qy 364 FAGCKKIFGSLAFLESPDGPASNTAPLOPEQLQVFTLEETIGYLYTSAWPDSLPLDS 423

Db 351 FVNCTKILGNLDELITGLNGDPWHKIPALDPEKLVNFRVTREITGYLVNTQSWPPHMNF 410

Qy 424 VFQNLQVIRGRILHNGAYS-LTQGLGISWLGRLSRELGLSLALIHNTLHCFVHTVPW 482

Db 411 VFSNLTITIGRSLYNRGFSLLIMKNLNTSLGFRSLKEISAGRIYISANRQLCYHHSLNW 470
Qy 483 DOLFERNHQALLHTA-NRPEDECVGEGGLACHQLCARGHCWGPGTQCVCNCSQFLRGQECV 541
Db 471 TKVLRGPTFEERLDIKHNPRDRCVAEGKVDPLCSSGCGWGPQGLCSRNRYGRGVGV 530
Qy 542 EECRVLOGLPREYVNAHCLCHPECCOPQNGSVTCFGEADQCVCACAHYKDPFPCVARCP 601
Db 531 THCNFLNGEPREFAHEACFSCHPECCOPMEGTATCNGSGSDTCAQCAHFRDGHPCVSSCP 590
Qy 602 SGVXPDLISYMPIWKPFBDEGACQCPINCTHSCVDLDDKGCPCABORA-----SPLTISUSA 657
Db 591 HGVLG--AKGPIKYKYPQVQNECRPCHENCHCTQCGKGPQLDGLGTLVLIGKTHLTALTY 648
Qy 658 VVGILLVVVLGVVGGILIKRQOKIR-KYTWRRLLQETELVEPLTPSGAMPNQAMRLK 716
Db 649 IAG--LVVIFWMLGGTFLYWRGRRIONKRAMRRYLERGESIEPLDPS-EKANKVLARIFK 705
Qy 717 EYIKANSKFTIGTEL-TVYKGIWIPDGENVKIPVAIKVLRENTSPKANKILDEAYMA 775
Db 706 ETELK-KLKVLSGSGVGTGKGVWIPGESIKIPVCIKVIEDKSGRSQFQAVTDHMLAIG 764
Qy 776 GVGSPYVRLGICLTSTVQLVTPQMPYVGLDHHVRENRLGSDLLNWCQIAKMSY 835
Db 765 SLDAHIVRLGLCPGSSQLVTVPLGSLDHHVRQHRGALGPQLLLNMGVQIAKMY 824
Qy 836 LEDVRLVHRDLAARNLVKSPNHVKITDPGLARLDDTEYHADGGKVPKMALESIL 895
Db 825 LEEHGMVHRNLAARNVLKSPQVADFGVADLLPPDKQLLYSEAKTPIKMALESIH 884
Qy 896 RRRFTHQSDVMSGYVTWELMTFGAKPYDGPAREIPDLLEKGERLPPOPTCTTIDVYIM 955
Db 885 FGKTHQSDVMSGYVTWELMTFGAEPYAGLAEVPLLEKGERLAQPOICTIDVYVM 944
Qy 956 VKCMWIDECRRFRELVSFSRWARDPQRFVVIQNEDLGPA---SPLDSTFYSLLEDD 1012
Db 945 VKCMWIDENIRTEKELANEFTRWARDPPRYLVIKRES-GPGIAPGEPHGLTKKLEEV 1003
Qy 1013 DMGDLVDAEEVLVPOOGFCDDPAPGAGMVHHRSSSTSGGDLTLGLEP-SBEEAP 1071
Db 1004 ELEPELDLDLEAED-----NLATTLGSAUSLPGTLNLRPGS 1044
Qy 1072 RSLAPSGAGSDVDFDGLMGAAGLQSLPTHQ-PSPLQRYSEDPVTPLP-----SET 1124
Db 1045 QSLSPSGGY-MPMNQNLGSCQESAVSGSSRCPRVSLH-----PMRPGCLASES 1097
Qy 1125 DGYVA-----PLTCSPOPE-----YVNPQDVRPPOPPSPREGP----- 1156
Db 1098 EGHVTSBAELQEKVSMCRSRSRSPRGRDSAYHSQRHSLLTPVTPLSPPGLEEDVN 1157
Qy 1157 ---LPAARPAGATLERAKTISP-GKNGVY-----KDVFAEGGAVENPEYLTPOGGAPO 1206
Db 1158 GYVMPDTHLKTGPSREGTLSSVLGSTEED-----EYENMNRRRHSP- 1208
Qy 1207 PHPPAPFAPDNLVYMD-----QDPERGAPPSTFKGTPTAENPEYL 1249
Db 1209 PHPPRPSLEBELGYEYMDVGDLSASLGSTQSCPLHPVIMPTAGTTPDEDIETM 1263

RESULT 10
JC4387
epidermal growth factor receptor homolog precursor - rat
N:Alternate names: ErbB3 protein; HER3 protein
C:Species: Rattus norvegicus (Norway rat)
C:Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C:Accession: JC4387
R:Hellyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A:Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.
A:Reference number: JC4387; MUID:96096535; PMID:8522190
A:Accession: JC4387
A:Molecule type: mRNA
A:Residues: 1-7339 <HEL>

A:Cross-references: GB:U29339; NID:G915389; PID:G915390
A:Experimental source: liver
A:Note: The authors translated the codon AAC for residue 369 as Thr and GTT for residue 370.
C:Comment: This protein is a functional heregulin receptor that transduces signals to the cell.
C:Genetics:
A:Gene: ErbB3
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homologous
C:Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F:640-659/Domain: transmembrane #status predicted <TM>
F:705-970/Domain: protein kinase homolog <KIN>
F:713-721/Region: protein kinase ATP-binding motif
F:939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr) (cc
Query Match 33.0%; Score 2246.5; DB 2; Length 1339;
Best Local Similarity 39.7%; Pred. No. 2.6e-85;
Matches 511; Conservative 173; Mismatches 441; Indels 161; Gaps 37;
Qy 3 LAALCRWGLLLALLPPGAA---STQVCTGTDMLRLPASPEHDLMLRHLHYQCGQVVOGN 59
Db 7 LQVLC-----FLLSLARGSEMGNSQAVCPGTLNGLSVTGDADNQYQTYLYKLYKECEVMGN 62
Qy 60 LETYLTPTNASLFLQDIOQEVGYVLIHNRQVPLQRLRIYVRGTQTFEDNVALAYLDN 119
Db 63 LEIVLTGHNADLSFLQWIREVTAYVLVAMNEFSLVPLNLRVVRGTQVYDGFKAIFVM-- 120
Qy 120 GDPNNTTPTVGASPGRLRLQRLSEITLKGVLTORNPOLCYQDTILWKDIFHKNNQL 179
Db 121 ---LNTNT-----NSSHALRQLKFTQTLSEILSGGVYIEKNDKLCMDTIDRDLVRVR-- 170
Qy 180 ALTLIDNRSRACHPCSPCKSGKRCWGESSEDCQSITRTVCAGGC-ARCKGPLTDCCHE 238
Db 171 GAIIVKNGANGCPCHEVCCKG-RCMGPGDDQIILTKIAPQCNCRGCPNPQCCHD 229
Qy 239 QCAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTTFESMPNPEGRYTFGASCVTA 298
Db 230 ECAGGCGSPQDTCFACRRFNDSGACVPRCPPELVVYNTKLTFOLEPNHTKYQYGGVCVAS 289
Qy 299 CPVNYLSTDVGSCTLVCPHLNQVTAEDGTORCEKSKPCARVCYGL--GMQIKANSKF 356
Db 290 CPNHFV-VDTQFCVRACPPDKMEVD-KHGLKMCPCGGLCPKACEGTSGSRYQTVDDSN 347
Qy 357 IGITELEFACCKIFGSLAFLPESFDGPASNTAPIQEOLOVFELEETITGLYLTISAWP 416
Db 348 ID-----GFVNCTKILGNLDFLTGLNVDPWHKIPALDPEKLVNFRVREITGLYNTQSWP 403
Qy 417 DSLPDLISVFONQIVIRGILHNGAYS-LTLQGLGISWGLSLRSLRSLGSLALIHNTLHC 475
Db 404 PMHNFVSFNLTTIGRSLYNRGFSLLIMKNLNTSLGFRSLKEISAGRIYISANQOLC 463
Qy 476 FVITVPWDQLFRPHQALLHTA-NRPEDECVGEGGLACHQLCARGHCWGPGTQCVCNCSQ 534
Db 464 YHSLNWTLLRGPSSEERLDIKYDRPLGELAEKGVCDPLCSSGCGWGPAGQGLSCRNY 523
Qy 535 LRQECVEECRVLOGLPREYVNAHCLCHPECCOPQNGSVTCFGEADQCVCACAHYKDP 594
Db 524 SRGVCVTHCNFLQGEPRFEVHEAQCFSCHEPCLEMTGISTYNGSGSDACARCAHFRDGP 583
Qy 595 FCVARCPGKVPDLISYMPIWKPFBDEGACQCPINCTHSC--VDLDDKGCPCABQASPLT 652
Db 584 HCYNSCPHTGLG-AKGPIKYKYPDAQNECHPCHECHCTQCGNCPQLDCLQAEVLSMKPH 641
Qy 653 SIYSVAVGILLVVVLGVVGGILIKRQOKIR-KYTWRRLLQETELVEPLTPSGAMPNQAO 711
Db 642 LVIAVTVG--LAVILMLGSGFLYWRGRRIONKRAMRRYLERGESIEPLDPS-EKANKVL 698
Qy 712 MRILKEQYIKANSKFTIGTEL-TVYKGIWIPDGENVKIPVAIKVLRENTSPKANKILDE 770
Db 699 ARIFKETELR-KUKVJGSGVGTGKGVWIPGESIKIPVCIKVIEDKSGRSQFQAVTDH 757
Qy 771 AYVMAGVSPYVSRLLGICLTSTVQLVTPQMPYVGLDHHVRENRLGSDLLNWCQIA 830

Query Match 23.8%; Score 1623; DB 1; Length 604;
Best Local Similarity 50.7%; Pred. No. 5.1e-60;
Matches 350; Conservative 76; Mismatches 137; Indels 128; Gaps 18;

QY 587 CAHYKDPFCVACPSGVKPDLSYMPIWKFPDEBAGACQPCPINCTHSCVDLDDKGCAPAEQ 646
DB 3 CAHFTDGHPCVKACPAVGENDTL-VKRYADANAVCQLCHPNCNCTRGKPGEGCP--- 58

QY 647 RASPLTSIVSAV-GILLVVVLGVVFGILLIKRQOKIRKYIMRRLLOSTELVEPLTPSGA 705
DB 59 NGSKTPSTAAGVVGGLLCLVVGIGLYLRRR-HIVAKRTLRRLLOSTELVEPLTPSGE 117

QY 706 MPNQMRILKEQYIKANSKFTIGTEL-TVYKGIWIPDGENVKIPVAIKVRENTSPKAN 764
DB 118 APNQAHLILKETEFK-KVKVLGSAFGTYKGLWIPESGKVPVAIKELREATSPKAN 176

QY 765 KEILDEAYMAGVGSPPYVRLIGLICTSTVQLTQMPYGCGLLDHVRNRRGLSQDILLN 824
DB 177 KEILDEAYMASVDNPHVCRLLIGLICTSTVQLTQMPYGCGLLDYIREHKNIGSQYLLN 236

QY 825 WCMQIAKMSYLEDLVLRDLAARNVLKSPNHVKITDFGLARLLDIDETEYHADGGKV 884
DB 237 WCVQIAKGMNYLEERLVRDLAARNVLKTPQHVKITDFGLAKLLGGADEKEYHAEGKV 296

QY 885 PIKMALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQ 944
DB 297 PIKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISSVLEKGERLPQ 356

QY 945 PICTIDVIMVWKMIIDSECRPRELVSFSEMRARDPQRFVVIQ-NEDLGPASPLDST 1003
DB 357 PICTIDVIMVWKMIIDSECRPRELVSFSEMRARDPQRFVVIQ-NEDLGPASPLDST 416

QY 1004 FYRSLLEDDMDLDAEYLVPOGFFCPDPAAGGAGGVHRRSSSTRSGGGDLTLGL 1063
DB 417 FYRSLLEDDMDLDAEYLVPOGFFCPDPAAGGAGGVHRRSSSTRSGGGDLTLGL 449

QY 1064 EPSEEEAPRSP-----APSEGAGSDVFDGOLGMAAGLQSLPHTDPSLQRYSEDPV 1118
DB 450 -----SRTPLLSLSATSNNSATNCID-----RNGQHPVREDTSFQRYSSDPTG 494

QY 1119 PLPST-DGVVAPITCSPQBYNVQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPG 1176
DB 495 NPLESIDDGLF-----PAPEYVQ--LMPKKPSTAM----- 524

QY 1177 KNGVVKVFAF-----GGAVENPEYLPQCGAAPQHPHPPAFSPAF 1217
DB 525 ---VQNYINFLSLTAISKLPMSRYQNSHSTAVDNPEYL-----NTNQSPLAKTVF 573

QY 1218 DNLVYWDQPPPERGAPPSTFKGTPTAENPEY 1248
DB 574 ESSPYWQSGNHQ-----INLDNPDY 594

RESULT 13
S35745
protein-tyrosine kinase (EC 2.7.1.112) erbb - avian erythroblastosis virus
C:Species: avian erythroblastosis virus
C>Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
C:Accession: S35745
R:Vennstrom, B.
submitted to the EMBL Data Library, March 1993
A:Reference number: S35743
A:Accession: S35745
A:Molecule type: DNA
A:Residues: 1-544 <VEN>
A:Cross-references: EMBL:X12707
C:Genetics:
A:Gene: erbb
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
F:133-400/Domain: protein kinase homology <KIN>
F:143-151/Region: protein kinase ATP-binding motif
F:170/Active site: Lys #status predicted

Query Match 23.1%; Score 1575; DB 2; Length 544;
Best Local Similarity 53.4%; Pred. No. 4.4e-58;
Matches 336; Conservative 70; Mismatches 129; Indels 94; Gaps 17;

QY 578 GPEADQCVACAHYKDPFCVACPSGVKPDLSYMPIWKFPDEBAGACQPCPINCTHSCVDL 637
DB 1 GP--DHCMKCAHFIDGPHCVKACPAVGENDTL-VMKYADANAVCQLCHPNCNCTRGKGP 57

QY 638 DDKGCAPAEQASPLTSIVSAV-GILLVVVLGVVFGILLIKRQOKIRKYIMRRLLOSTEL 696
DB 58 GLEGCP---NGSKTPSTAAGVVGGLLCLVVGIGLYLRRR-HIVAKRTLRRLLOSTEL 113

QY 697 VEPLTPSGAMPNOAMRILKEQYIKANSKFTIGTEL-TVYKGIWIPDGENVKIPVAIKVL 755
DB 114 VEPLTPSGEAPNQAHLILKETEFK-KVKVLGSAFGTYKGLWIPESGKVPVAIKEL 172

QY 756 RENTSPKANKEILDEAYMAGVGSPPYVRLIGLICTSTVQLTQMPYGCGLLDHVRNRRG 815
DB 173 REATSPKANKEILDEAYMASVDNPHVCRLLIGLICTSTVQLTQMPYGCGLLDYIREHKD 232

QY 816 RLGSODLLNWMQIAKMSYLEDLVLRDLAARNVLKSPNHVKITDFGLARLLDIDET 875
DB 233 NIGSQYLLNWCQIAKGMNYLEERLVRDLAARNVLKTPQHVKITDFGLAKQLGADEK 292

QY 876 EYHADGGKVPIKMALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDL 935
DB 293 EYHADGGKVPIKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISSV 352

QY 936 EKERLPQPPICTIDVIMVWKMIIDSECRPRELVSFSEMRARDPQRFVVIQ-NEDL 994
DB 353 EKERLPQPPICTIDVIMVWKMIIDSECRPRELVSFSEMRARDPQRFVVIQ-NEDL 412

QY 995 GPASPLDSTYRSLLEDDMDLDAEYLVPOGFFCPDPAAGGAGGVHRRSSSTRS 1054
DB 413 HLPSPSTDSKYRSLLEDDMDLDAEYLVPOGFFCPDPAAGGAGGVHRRSSSTRS 454

QY 1055 GGGDLTLGLPSEEEAPRSP-----APSEGAGSDVFDGOLGMAAGLQSLPHTDPSPL 1109
DB 455 -----SRTPLLSLSATSNNSATNCIDRNGG-----H----- 481

QY 1110 QRYSEDPVLPSETDGVVAPITCSPQBYNVQDVRPQPPSPREGPLPAARPAAGAT-LE 1168
DB 482 -----PVREDGFL-----PAPEYVQ--LMPKKPSTAMVQNYINFLSLTAIS 522

QY 1169 RAKTLSPGKNGVVKVFAFGAVENPEYL 1197
DB 523 KLPIDSRYQN-----SHSTAVDNPEYL 544

RESULT 14
S00727
kinase-related transforming protein (erbB) (EC 2.7.1.1) - avian erythroblastosis virus
C:Species: avian erythroblastosis virus
C>Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
C:Accession: S00727
R:Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
Oncogene Res. 1, 265-278, 1987
A:Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mutant
A:Reference number: S00727; MUID:88217326; PMID:2897102
A:Accession: S00727
A:Molecule type: DNA
A:Residues: 1-545 <SCO>
A:Cross-references: EMBL:X06943
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; phosphotransferase
F:135-400/Domain: protein kinase homology <KIN>
F:143-151/Region: protein kinase ATP-binding motif

Query Match 23.0%; Score 1568; DB 2; Length 545;
Best Local Similarity 53.4%; Pred. No. 8.5e-58;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:10:23 ; Search time 22.179 Seconds
(without alignments)
5413.772 Million cell updates/sec

Title: SEQ4-710-730-12

Perfect score: 6775

Sequence: 1 MEAAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1249

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 73.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6602	97.4	1255	1 A2471	protein-tyrosine k
2	5805	85.7	1260	1 TVRTNU	protein-tyrosine k
3	5795.5	85.5	1254	2 I48151	p-185 precursor -
4	3030	44.7	1210	1 GQHUE	epidermal growth f
5	3001	44.3	1210	2 A53183	epidermal growth f
6	2978.5	44.0	1223	1 TVCHIV	epidermal growth f
7	2849.5	42.1	1308	2 A47253	epidermal growth f
8	2553	37.7	1166	1 S06142	protein-tyrosine k
9	2305.5	34.0	1342	2 A36223	kinase-related tra
10	2218.5	32.7	1339	2 JC4387	epidermal growth f
11	1650.5	24.4	698	1 TVFVLV	protein-tyrosine k
12	1588	23.4	604	1 TVYUHV	protein-tyrosine k
13	1537	22.7	544	2 S35745	protein-tyrosine k
14	1530	22.6	545	2 S00727	kinase-related tra
15	1523.5	22.5	1330	1 GQFFE	epidermal growth f
16	1513	22.3	540	2 B44776	protein-tyrosine k
17	1511	22.3	540	1 TVFVBB	protein-tyrosine k
18	1509	22.3	644	2 A36325	epidermal growth f
19	1224	18.1	1323	2 E88257	protein let-23 lim
20	1224	18.1	1374	2 S70712	protein-tyrosine k
21	1139	16.8	1369	2 S70713	protein-tyrosine k
22	1126	16.6	527	2 A40232	protein-tyrosine k
23	1109	16.4	1717	1 A45558	epidermal growth f
24	975.5	14.4	843	2 A27131	epidermal growth f
25	806.5	11.9	346	2 S13807	protein-tyrosine k
26	754.5	11.1	311	2 S13808	protein-tyrosine k
27	703	10.4	1363	2 T43220	insulin-like growth
28	676.5	10.0	1372	2 A34157	insulin receptor p
29	675	10.0	1382	1 INHUR	insulin receptor p

ALIGNMENTS

RESULT 1

A24571

protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N/Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein erbB2

C/Species: Homo sapiens (man)

C/Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999

C/Accession: A24571; A25491; A44188; B44188; I59509; I57622

R/Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;

Nature 319, 230-234, 1986

A/Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth factor

A/Reference number: A24571; MUID:86118663; PMID:3003577

A/Accession: A24571

A/Molecule type: mRNA

A/Residues: 1-1255 <YAM>

R/Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.

Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A/Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epidermal

A/Reference number: A25491; MUID:86016729; PMID:2995967

A/Accession: A25491

A/Molecule type: DNA

A/Residues: 737-1031 <SEM>

A/Cross-references: GB:M11767; NID:G182163; PIDN:AAA35808.1; PID:G553282

R/Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg, P.

Science 230, 1132-1139, 1985

A/Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal

A/Reference number: A44188; MUID:86070181; PMID:2999974

A/Accession: A44188

A/Molecule type: DNA

A/Residues: 740-910 <COU1>

A/Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989

A/Accession: B44188

A/Molecule type: mRNA

A/Residues: 1-517, 'FALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>

A/Cross-references: GB:M11730; NID:G183986

R/King, C.R.; Kraus, M.H.; Aaronson, S.A.

Science 229, 974-976, 1985

A/Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A/Reference number: I59509; MUID:85272597; PMID:2992089

A/Accession: I59509

A/Status: translated from GB/EMBL/DDBJ

A/Molecule type: DNA

A/Residues: 832-909 <REX>

A/Cross-references: GB:I29395; NID:G459807; PIDN:AAA35809.1; PID:G459808

R/Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987

A/Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptional

A/Reference number: I57622; MUID:87286898; PMID:3039351

A/Accession: I57622

A/Status: translated from GB/EMBL/DDBJ

A/Molecule type: DNA

A/Residues: 1-191 <TAL>

A;Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
 C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
 C;Genetics:
 A;Gene: GDB:ERBB2; NGU; NEU; HER-2
 A;Cross-references: GDB:120613; OMIM:164870
 A;Map position: 17q21.1-17q21.1
 A;Introns: 25/1; 75/3; 147/1; 983/3
 A;Note: the list of introns is incomplete
 C;Function:
 A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
 C;Superfamily: epidermal growth factor receptor; protein kinase homology
 C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
 inase
 F;1-21/Domain: signal sequence #status predicted <SIG>
 F;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
 F;22-653/Domain: extracellular #status predicted <EXT>
 F;70-304/Domain: EGF receptor extracellular domain repeat <E81>
 F;395-605/Domain: EGF receptor extracellular domain repeat <BE2>
 F;654-675/Domain: transmembrane #status predicted <TMM>
 F;676-1255/Domain: intracellular #status predicted <INT>
 F;718-983/Domain: protein kinase homology <KIN>
 F;726-734/Region: protein kinase ATP-binding motif
 F;68,124,187,259,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
 F;753/Active site: Lys #status predicted
 F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 97.4%; Score 6602; DB 1; Length 1255;
 Best Local Similarity 97.2%; Pred. No. 3.2e-277;
 Matches 1221; Conservative 6; Mismatches 21; Indels 8; Gaps 2;

Qy 1 MELAALCRWGLLLALLPPGAASCTGTDKMLRASPETHDMLRHLVQGCQVVOGNL 60
 Db 1 MELAALCRWGLLLALLPPGAASCTGTDKMLRASPETHDMLRHLVQGCQVVOGNL 60

Qy 61 ELTYLPTNASLSFLQDIQEVGYVLIHQNVRQVPLQRLRVRGTLQFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLSFLQDIQEVGYVLIHQNVRQVPLQRLRVRGTLQFEDNYALAVLDNG 120

Qy 121 DPLNTTTPVTGASPGGLREQLRLSLTEILKGVLIQNPOLCYQDTLLKDKDIFHNKOLA 180
 Db 121 DPLNTTTPVTGASPGGLREQLRLSLTEILKGVLIQNPOLCYQDTLLKDKDIFHNKOLA 180

Qy 181 LTLIDTNRSRACHPCSPMKSGSRGWGSSDQSLRTVACGGCARCKGLPTDCHEQC 240
 Db 181 LTLIDTNRSRACHPCSPMKSGSRGWGSSDQSLRTVACGGCARCKGLPTDCHEQC 240

Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300
 Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300

Qy 301 YNYLSTVGSTLYCPLHNQEVTAEDGTQCEKSKPCARVCYGLGHWYIKANSKFTGIT 360
 Db 301 YNYLSTVGSTLYCPLHNQEVTAEDGTQCEKSKPCARVCYGLGHWYIKANSKFTGIT 360

Qy 361 ELEFAGCKKIFGSLAFIPESDGPASNTAPLQEQLOVFETLEITGYLYISAWPDSLP 420
 Db 361 IQEFAGCKKIFGSLAFIPESDGPASNTAPLQEQLOVFETLEITGYLYISAWPDSLP 420

Qy 421 DLSVFQNLQVIRGRIHLNGAYSLTQGLIGISWLGLRLSRLREIGSLALIHNTLHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRIHLNGAYSLTQGLIGISWLGLRLSRLREIGSLALIHNTLHLCFVHTV 480

Qy 481 PWDOLFNPQHALLHTANRDEDECVGGLACHQICARGHCWGPGPTQCVNCSQFLRQEC 540
 Db 481 PWDOLFNPQHALLHTANRDEDECVGGLACHQICARGHCWGPGPTQCVNCSQFLRQEC 540

Qy 541 VEECRLVQLGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQCVACAHYKDPFCVARC 600
 Db 541 VEECRLVQLGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQCVACAHYKDPFCVARC 600

Qy 601 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSTISAVVG 660
 Db 601 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSTISAVVG 660

Db 601 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSTISAVVG 660
 Qy 661 ILVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRLKTEL 720
 Db 661 ILVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRLKTEL 720
 Qy 721 RKVKVLGSGAGFGQYIKA-----NSKFIGITELVIRENTSPKANKEILDEAYVMAGVGS 773
 Db 721 RKVKVLGSGAGFGTVYGIWIPDGENVK-IPVAIKVIRENTSPKANKEILDEAYVMAGVGS 779
 Qy 774 PYVSRLIGICLTSTVQLVTQLMYPYGLLDHVRNRRGLSQDILLNWCQIAGKMSYLEDV 833
 Db 780 PYVSRLIGICLTSTVQLVTQLMYPYGLLDHVRNRRGLSQDILLNWCQIAGKMSYLEDV 839
 Qy 834 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKMMALRESILRRRF 893
 Db 840 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKMMALRESILRRRF 899
 Qy 894 THQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMWKWC 953
 Db 900 THQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMWKWC 959
 Qy 954 MIDSECRPRPRELVSFSPMRARDPQRFVJTQNEIDLGPASPLDSTFYRSLLEDDDDMGDLVD 1013
 Db 960 MIDSECRPRPRELVSFSPMRARDPQRFVJTQNEIDLGPASPLDSTFYRSLLEDDDDMGDLVD 1019
 Qy 1014 AEYLVPOQGFCCPDPAAGAGGMVHRHRSSTRSGGDLTLGLESEEEAPRSLAPSE 1073
 Db 1020 AEYLVPOQGFCCPDPAAGAGGMVHRHRSSTRSGGDLTLGLESEEEAPRSLAPSE 1079
 Qy 1074 GAGSDVDFDGLGMAAKGLQSLTFHDPSPLOQRYSEDPVPLPSETDGYVAPLTCSPQPEY 1133
 Db 1080 GAGSDVDFDGLGMAAKGLQSLTFHDPSPLOQRYSEDPVPLPSETDGYVAPLTCSPQPEY 1139
 Qy 1134 VNQPDVRRPQPPSPREGPLPAARPAATLERAKTSLSPGKGVKVDVFAFGGAVENPEYLTTP 1193
 Db 1140 VNQPDVRRPQPPSPREGPLPAARPAATLERAKTSLSPGKGVKVDVFAFGGAVENPEYLTTP 1199
 Qy 1194 QGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1249
 Db 1200 QGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255

RESULT 2

TVRTNU

protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999

C;Accession: A24562; A61204

R;Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.

Nature 319, 226-230, 1986

A;Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.

A;Reference number: A24562; MUID:86118662; PMID:3945311

A;Accession: A24562

A;Molecule type: mRNA

A;Residues: 1-1260 <BAR>

A;Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746

R;Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen,

Carcinogenesis 12, 1975-1978, 1991

A;Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no m

2-thiazolyl]formamide or N-methyl-N-nitrosourea.

A;Reference number: A61204; MUID:92035293; PMID:1682063

A;Accession: A61204

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 637-663,'V',665-702 <MAS>

A;Note: authors translated the codon GCA for residue 25 as Val

C;Genetics:

A;Gene: neu

C;Superfamily: epidermal growth factor receptor; protein kinase homology

C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

A;Molecule type: mRNA
 A;Residues: 969-971,'K',973-1115,'D' <EIS>
 A;Cross-references: EMBL:Z12608
 R;Heisermann, G.J.; Gill, G.N.
 J. Biol. Chem. 263, 13152-13158, 1988
 A;Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
 A;Reference number: A28941; MUID:88330814; PMID:3138233
 A;Accession: A28941
 A;Molecule type: protein
 A;Residues: 689-694,'X',696-704,'L',706-707,989-992,'XX',995-996,'X',998-1000;1002-1009,
 R;Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
 submitted to the EMBL Data Library, April 1994
 A;Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
 A;Reference number: S45325
 A;Accession: S45325
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-971,'K',973-1210 <VER>
 A;Cross-references: EMBL:X78987; NID:G488830; PIDN:CAA55587.1; PID:G488831
 R;Paria, B.C.; Das, S.K.; Andrews, G.K.; Dev, S.K.
 Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
 A;Title: Expression of the epidermal growth factor receptor gene is regulated in mouse b
 A;Reference number: I49643; MUID:93126380; PMID:7678348
 A;Accession: I49643
 A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 12-20,132 <RES>
 A;Cross-references: GB:I06864; NID:G193001; PIDN:AAAS3029.1; PID:G567201
 C;Genetics:
 C;Superfamily: epidermal growth factor receptor; protein kinase homology
 C;Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop
 F;1-24/Domain: signal sequence #status predicted <SIG>
 F;648-670/Domain: transmembrane #status predicted <TM>
 F;712-977/Domain: protein kinase homology <Kin>
 F;680,695/Binding site: phosphate (Thr) (covalent) #status experimental
 F;697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
 F;993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
 F;1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
 F;1197/Binding site: phosphate (Tyr) (covalent) #status experimental
 Query Match 44.3%; Score 3001; DB 2; Length 1210;
 Best Local Similarity 48.6%; Pred. No. 3.8e-122;
 Matches 620; Conservative 168; Mismatches 365; Indels 122; Gaps 28;
 11 LLALALLPGAA--STQCTGTDKMLRLPASPETHLMLRLHYOGCQVQGNLEILYPTN 68
 14 LLTALCAAGALBEKKYCCQTSNRLTQGTGTFEDHFLSLQRMNCEVTLGNLEITYQRN 73
 69 ASLSFLQDIOEVGYVLIHANOVRQVPLQRLRIVRGTQOLFEDNYVALVLDGDPNNTTP 128
 74 YDLSFLTKTIOEVAGYVLIANTVERIPLENLQIIRGNALYENTVALAISN----- 124
 129 VTGASPGGLRELQRLSITELKGGVLTQRNPOLCYODTILWKDI----FHKNNQLAULTI 184
 125 -YGNRTGLRELPMNLQELILGAVRFNNPILCNMDTIQMRDVIQNVFNWSMDL--- 180
 185 DTVRSRACHPCSPMKSGRSRWGSSSEDCQLSTRITVACGGCA-RCKGLPTDCHQCAAG 243
 181 -QSHSSCPKCDPSCPNCSGNGSCWGGEENCQKLTIIIAQQCSHRCRCRSPSDCHNCAAG 239
 244 CTGPKHSDDLACLFHNHSGICELHCPALVTYNTDTFESMENPEGRYTFGASCYTACPNY 303
 240 CTGPRESDCLVCKQFQDEATCKDTCPPLMLNPTTQMVNPNPKGYSFGATCVKKCPNY 299
 304 LSTDVSGCTLVCLPHNQEVTAEDGTORCEKSCPKARVCYGLQWQIKANSKFIGITELE 363
 ..300 VVTDHSCVRACGPDYIEV-EEDGIRKCKCDGCRKVCNGIGIGEFK-DTLSINATNIK 357
 364 -FAGCKIFGSLAFLESFDPGDPASNTAPLQPSQLQVFTLEITGYLIYSANWPSLDL 422
 358 HFKYCTAISGDLHLPLVAFKGSDFTRTPPLDPRLEILTKVKEITGFLTAQWPDNWTDL 417

RESULT 6
TVCHLV

epidermal growth factor receptor precursor - chicken
 N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
 C;Species: Gallus gallus (chicken)
 C;Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
 C;Accession: A27720; A00643
 R;Lax, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennstr
 Mol. Cell. Biol. 8, 1970-1978, 1988
 A;Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou
 A;Reference number: A27720; MUID:88261272; PMID:3260329

423 SVFQNLQVIRGRILHNGAYSLTQGLGISWLGRLSRELGSGLALIHNTLHLCFVHTVPW 482
 418 HAFENLEIIRGRTRKQGFSLAVGLNITSLGLRSKEISDGVIIISGRNLCYANTINW 477
 483 DQLFNPHQALLHTANRDECEVGEGLAQHLCARGHCWGPQTQCNCQSQFLRGECVE 542
 478 KKLFGTPNQTKIMNNAEKDCKAVNHVCNPLCSSEGCGWPEPRDCVSCQNVSRGECVE 537
 543 ECEVLOGLPREYVNAHCHLPCHPECOPONGSVTCFGEADQCVACAHYKDPPECVACRPS 602
 538 KCNILEGEPREFVENSECQCHPECLPQAWNITCTGRGPDNCIQCAHYIDGPHCVKTCPA 597
 603 GVKPDLNYPWPKFPEDEGACQPCPNTCHSCVDLDDKGPABQBRASPLTISVAVVGIL 662
 598 GINGENNTL-VMKYADANNVCHLCHANCYGCAGPGLQCEVWPSGPKPSIATIGVGL 656
 663 LVVLGVVFGI-LIKRQOKIRKYWRRLQETELVEPLTPSGAMPNQAMRLIKTELRL 721
 657 LFTIW-VALGIGLFMRERRHVRKRTLRLLQERELVEPLTPSGEAPNQAHRLIKETEFK 715
 722 KYVLGSGAFGQ-----YIKANSKF---IGITELVRENTSPKANKEIILDEAYVMAGVS 773
 716 KIKVLGSGAFGVYKGLWIPGEKVKIPVAIKE--LREATSPKANKEIILDEAYVMASVDN 773
 774 PYVSRLLGICLTSTVOLTPYGLLDHVRNRCRLGSDLLNWCMIQAKGMSYLEDV 833
 774 PHVCRLLGICLTSTVOLTPYGLLDYVREHKDNIGSYLLNWCVIQAKGMVLEDR 833
 834 RLVRDLAARNVLKSPNPKVITDFGLARLLDIDETEHADGKVPKIMMALESILRRRF 893
 834 RLVRDLAARNVLKTPQHKITDFGLAKLGAEEKEYHAEGGKVPKIMMALESILHRIY 893
 894 THQSDVMSYGVTVWELMTGAKPYDGPAREPDLLEKGERLPDPPICTIDVTYMWVKCW 953
 894 THQSDVMSYGVTVWELMTGSKPYDGPASDISILEKGERLPDPPICTIDVTYMWVKCW 953
 954 MIDSECRPFRELVSFSESMARDPQRFVIO-NEDLGASPLDSTFYRSLLEDDMDGLV 1012
 954 MIDADSRPFRELILEFSQWARDPQRYLVIOGDERHPLSPDTSNFYRALMDEDMEDVV 1013
 1013 DABEYLVPOQGFPCDPAPGAGGMVHRRHSSTRSGGDLTLGLPESEEAAPRSLAPS 1072
 1014 DADYELIPQGGF-----NSPST-----SRTPLSS 1039
 1073 EGNAGSDVFDGLMGAAKGLQSLPHDPSLQYSESDTVPPLPSET--DGYVAPLTCSPQ 1130
 1040 LSATSN-----NSTVACINRNGSCRVKEDAFQRYSSDPTGCAVTEDNIDDAFL-----PV 1089
 1131 PEYVNPDPVYRPPSPREGPLPAARPAAGATLERAKTLPCKNGVVKDVPAGGAVENPEY 1190
 1090 PEYVNO-SYPKRPAQSVQNPVYHNPQLHP-----AFORDLHYQN--PHSNVAGNPEY 1138
 1191 L-TPQGAAPHPPPPAPFADNLYWDQ-----DP-----PERGAPPSTFKG 1233
 1139 LNTAQ-----PTCLSSGFSNPAWIKGSHQMSLDNPDYQODFFPKETKPNKIFKG 1189
 1234 TPTAENPEYLGLDVP 1248
 1190 -PTAENAEYLRVAPP 1203

A;Accession: A27720
A;Molecule type: mRNA
A;Residues: 1-1223 <LAX>
A;Cross-references: GB:M20386
R;Nilsson, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1995
A;Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and product formation
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: A00643
A;Molecule type: mRNA
A;Residues: 585-1223 <NIL>
A;Cross-references: GB:M10066
C;Genetics:
A;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor specific protein kinase
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F;31-654/Domain: extracellular #status predicted <EXT>
F;81-307/Domain: EGF receptor extracellular domain repeat <EE1>
F;397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F;655-677/Domain: transmembrane #status predicted <TM>
F;678-1223/Domain: intracellular #status predicted <INT>
F;719-984/Domain: protein kinase homology <KIN>
F;727-735/Region: protein kinase ATP-binding motif
F;136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #
F;192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F;687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;754/Active site: Lys #status predicted
F;1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predicted
Query Match 44.0%; Score 2978.5; DB 1; Length 1223;
Best Local Similarity 47.5%; Pred. No. 3.6e-121;
Matches 618; Conservative 173; Mismatches 352; Indels 157; Gaps 30;
QY 8 RWGLLLALLPPGAA-----STOVCTGTMKRLPASPETHLMLRHLHYGGCVQVGNLE 61
DB 13 RGAALVLLLLGVALCSAVEBEKVCQCTNNKLTQLGHVEDHFTSLQRMVNNCEVLSNLE 72
QY 62 LTYLPTNASIFLODIOEVGYVLIHNOVQVPLQRLRVGTQTFQEDNYALAVLNDG 121
DB 73 ITVEHNRDLTFUKTIOEVAGYVLIHNMVDVPLENLQIIRGNVLYDVSFALAVLSNYH 132
QY 122 PLNNTTPTVTGASPGGLRELQRLSLTEILKGGLVLIQNPOLCYQDTILWKDIFHKNQAL 181
DB 133 -NNKTO-----GLRELPMKRLSELINGVKVLSNNPKLNMVTLVNDIITSRK-PL 182
QY 182 TLID-TNRSRACHPCSPMKSGKSCWGESSEDCQSLRTVTCAGCA-RCKGFLPTDCHEQ 239
DB 183 TVLDFASNLSSCPKCHPNCTEDHCWAGEQNCQTLTKVICAQCSGRCGRKVPSCCHNQ 242
QY 240 CAACTGPKHSDCLACLNHNSGICELHCPALVYNTDTFESMPNPEGRVTCGASCVTAC 299
DB 243 CAACTGTPRESDCCLACRFRDDATCKDTCPLLYNPTTYQMDVNPGEKYSFGATCVR 302
QY 300 PYNLYSTDVSGCTVCPPLHNOEYTAEDGTORCEKCKPCARVCYGLGMQVYKANSKFIGI 359
DB 303 PHNVVTDHGSCVRSNNTDYEV-ENGVRKCKCDGLSKVCNGIGELKGLS-INA 360
QY 360 TELE-PAGCKKIFGSLAFIPESFDGDPASNTAPLOPEQLQVFTLEITGYLYISAMPDS 418
DB 361 TNIDSPKNTCKINGDVSILPVAFGLDAFTKTLPLDKKLDVFTVKEISGFLLIQAWPDN 420
QY 419 LPDLSVFONLOVTRGILHNGAVSLTLOGISWLGRLSRLGSLGLALIHNTLHCFVH 478
DB 421 ATDLYAFENLEIIRGTRKQGVSLAVNLKIOSLGLRSLSKETSDDGAIAMKNKLCYAD 480
QY 479 TVPWDOLFRRPHQALLHTANRPECVGEGLACHQLCARGHCWGPPTQCVNCSQFLRGQ 538
DB 481 TNNWRSLFATQSKTKLIQNRKNKNDCTADRHVCDPLCSGVCGWGPFPFHCFSRFFSRQ 540
QY 539 ECVECRVLQGLPREYVNAHCLPCHPECPQNG---SVTCFGPEADQCACAHYKDPFF 595

DB 541 ECVKQCNILQGEPRFERDSKCLPCHSECLVQNSTAYNTTCSGPGDHCMKCAHFIDGPH 600
QY 596 CVARCPGKPDLSYMPINWKFDPDEECACOPCPINCTHSCVDLDDKGCPCABQASPLTSIV 655
DB 601 CVKACFAGVLGENDTL-VWKYDANAVCOLCHPNCTRGCKGPGLEGP---NGSKTPSIA 656
QY 656 SAVV-GILLVWVGVFGILIKRQOKIRKYNNRLLQETELVEPLTPSGAMPNQAMRI 714
DB 657 AGVVGGLLCLVVGVLGILGYLRRR-HVVRKTLRRLQLERELVEPLTPSGEAPNQAHRI 715
QY 715 LKTELKRVKVLGSGAFQ-----YIKANSKP---IGITELVLRNTSPKANKEILDEAY 766
DB 716 LKETEKKVKVLGSGAFGVYKGLWPEGEKVIPAIRE--LREATSPKANKEILDEAY 773
QY 767 VMAGVSPVSVRLGLCLTSTVQLMPYGLLDHVRNCRGLSGQDLNNCMQIAGK 826
DB 774 VMASVDPHVCRLLGLCLTSTVQLMPYGLLDYIREHKONIGSOYLLNMCVQIAGK 833
QY 827 MSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMAL 886
DB 834 MNYLEERLVRDLAARNVLKTPQHVKITDFGLAKLGADEKEYHAEGGKVPKWMAL 893
QY 887 SILRRFTQSDVWSYGVTVWELMTFGAKYDGIPIAREIPDLLEKGERLPPOPICTIDVY 946
DB 894 SILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPIASEISSVLEKGERLPPOPICTIDVY 953
QY 947 MIMVKCMMIDSECRPRELVESEFSRMARDPQRFVIO-NEDLGPAASPLDSTFYRSIL 1005
DB 954 MIMVKCMMIDUSRPFRELIAEFKWARDPPRYLVIOQDERMHLSPDTSKIFYRTUMEE 1013
QY 1006 DMGDLVDAAEYLVPOQGFPCDPAPGAGMVHRRHRSSTSSGGDLTLGLPSEBEAP 1065
DB 1014 EDMEDIVDADEYLVPHQGF-----NSPST-----S 1039
QY 1066 RSPL-----APSEGAGSDVFDGLGMAAKGQSLTHDPSPLQRYSEPTVPLPSET-- 1118
DB 1040 RTPLLSSLSATSNNSATNCID-----RNGQGHVREDSFVQRYSSDPTGNFLESID 1091
QY 1119 DGVVAPLTCFPOEYVNPDPVPQPSREGPLPAARPAAGATLERAKTLSPGNGVVKOV 1178
DB 1092 DGLF-----FAPEYVNO--LMPKRES-----TAMVQONLYNNI 1122
QY 1179 F-----AFGAVENPEYLTPOGGAAPQHPPPAFSPAFNLYYWDQ---- 1219
DB 1123 SLTAISKLPWDSYQNSHSTAVDNPEYL-----NTNQSLAKTVFESSPTWIOSGNH 1174
QY 1220 ----DPPE-----RGAPPSTFKGPTPTAENPEYLGLDVP 1248
DB 1175 QINLONPDYQQOFLPNETKPNGLLKVPAAENPEYLRVAAP 1214

RESULT 7

A47253

Epidermal growth factor receptor, HER4 - human

C;Species: Homo sapiens (man)

C;Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999

C;Accession: A47253

R;Flouman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; Ne

Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993

A;Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal

A;Reference number: A47253; MUID:93189574; PMID:8383326

A;Accession: A47253

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: nucleic acid

A;Residues: 1-1308 <PLO>

A;Cross-references: GB:L07868; NID:G37359; PIDN:AAB59446.1; PID:G37360

A;Note: sequence extracted from NCBI backbone (NCBI:126842)

C;Superfamily: epidermal growth factor receptor; protein kinase homology

C;Keywords: ATP; growth factor receptor

F;716-981/Domain: protein kinase homology <KIN>

F;724-732/Region: protein kinase ATP-binding motif

Query Match		42.1%	Score 2849.5;	DB 2;	Length 1308;
Best Local Similarity		44.1%;	Pred. No. 1.4e-115;		
Matches		593;	Conservative 184;	Mismatches 390;	Indels 179; Gaps 29;
QY	9	WGILLALLPGAA-----	STOVCTGTDMMKRLRASPETHDMLRHLHYQGCVQVQGNLELTY	64	
Db	8	WVWVSLVAAAGTVQPSDSQSVAGTENKLSLSLEQQYRALRYKYYENCEVWGNLEITS	67		
QY	65	LPTNASISFLQDIOEVGYVLIHNOVROVQLORLIVRGTOLPEDNYVALAVLNDGDLN	124		
Db	68	IEHRDLSURKREVTGVYVVALNQRYLPLENRLRIIRGTGLYEDYALAIENVRKOG	127		
QY	125	NTTPVTGASPGGLRELQRLSITELKGGVLIQIORNPOLCYQDTILWKDIFHKNQLALTLI	184		
Db	128	NF-----GLQELGLKNLTELINGVYVDQNKFLCYADTIHWQDIVERNPWSNLTIV	178		
QY	185	DYNRSRACHPCSPMKSGSRGWSSSEDCQSLTRTVACGGC-ARCKGLPLPTDCCHEQCAAG	243		
Db	179	STNGSSCGGRCHKSCG-RCWGPTENHCQTLTRTVCAEQCDGRCYGYPVSDCCCHRECAAG	237		
QY	244	CTGPKSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCYVACPYNV	303		
Db	238	CSGPKDTCFACMFNDGACVTCQPTFYVNPFTTFQLEHFNKAYTYGAFCKKCPHNF	297		
QY	304	LSTDVGSCTLVCPLNHOEVAEDGTQRCCKSPCARVCYGLGMQYIKANSKFTIGITELE	363		
Db	298	V-VDSSCVACPSKMEV--EENGIMCKEPTDICPKACDGTGSLMSAQTVDSNIDK	355		
QY	364	PAGCKIFGSLAFESFDGDPASNTAPLOEQVLFETLEETGYLYISAWPDSLPDLS	423		
Db	356	FINCTKINGNLIFLVTGIHGDYNAIEAIDPEKLNVERTVREITGFLNIQSWPNNMTDFS	415		
QY	424	VFNQLOVIRGILHNGAYSITLQGLGISWGLSRLGSLALIHNTHLCFVHTVPWD	483		
Db	416	VFNLVTVIGRVLVYGLSLILKQGGITSLQFOSLKEISAGNTYITDNLNLCYHTINWT	475		
QY	484	QLFRNPHQALLHTANRDEBCVGEGLACHOLCARGHCWGPQTQVCNCSQFLRGQECVEE	543		
Db	476	TLFSTINQIRVIRDNRAENKAENCAEAGMVCNHLCSDDGCGPQDCLSCRFSGRICIES	535		
QY	544	CRVLOGLPRYVNAHCLPCHBQCP--ONGSVTCFGPEADQCVACAHYKDPDPPFCVACRPS	602		
Db	536	CNLYDGEFREFFENGISICEVCDPQCEKMEGDLTCHGPGDNCCTKSHFKDGNCKVEKCPD	595		
QY	603	GVKPDLSYMIWPFPEEGACQCPNTCHSDVDLDDKGC-----PAEQRASPL	651		
Db	596	GLOQANSF--IFKYADPDRECHPCPNCTGCGNGPTSHDCIYVPTWGHSTLPQHAR--TPL	652		
QY	652	TSIVSAVV--GILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNOA	710		
Db	653	--TAAGVIGLFLIVLIGLTFAYVVRKSIK--KRALRRFL--ETELVEPLTPSGTAPNOA	708		
QY	711	QMILKETELRKVKVLSGAFQYIK-----ANSKFIGITELVLRNTSPKANKELDE	764		
Db	709	QLRILKETELRKVKVLSGAFQYIKGIWVPEGETVKPIVAIKILNETTGPKANVEFMDE	768		
QY	765	AYWAGVGSVYVSRLLGICLTSTVOLVTLMPYGCILLDHYENRGRGLSGDILNWCQIA	824		
Db	769	ALIMASMDHPLHVLGVCLSPITQLVTLMPHGCLLEYVHEHKDNIGSOLLNWCQIA	828		
QY	825	KGMSYLEDVRLVHRDLAARNVLKSPNHVKITDIFGLARLLDIDETEYHADGKVPKQMA	884		
Db	829	KGMVLEERLVRDLAARNVLKSPNHVKITDIFGLARLLGEBEKEYNADGGKMPKQMA	888		
QY	885	LESILRRFTHQSDVWSYGVVWELMTFGAKYDGPAPRIIPDLLEKGERLPPOPICTID	944		
Db	889	LECIVHKKFTHQSDVWSYGVVWELMTFGKPKDGPITREIPDLLEKGERLPPOPICTID	948		
QY	945	VYIMVKKMIDSECPREFELVSESRMARDPQFVVIQNEP-LGPASPLDSTFYRSL	1003		
Db	949	VYIMVKKMIDADSRPKFELAAEFSESRMARDPQFVLIQGGDRMKLPSPNDSKFFQNL	1008		
QY	1004	EDDDMGDLVDABEYLPQOQFFCPDPAPGAGGVHHRHRSSTRSGGDLTLGLEPSEEE	1063		

Db	1009	DEEDLEDMDABEYLP-QAFNIPP-----IYTSRARIDSNRS-----EIGHSPPPAY	1056		
QY	1064	APRS-----PLAP-SEGAGSDVFDGDLGMAAKGLOS	1094		
Db	1057	TPMSGNQFYVRDGGFAAGQSVVPYRAPRTSTIPEAFVAQGATAEIFFDDSCCNGTLRKPVA	1116		
QY	1095	LPTHDPSPLOQYSEDPTVPLPS-----ETDCGYVAPLTCSPQEPYVNOVDVRPQPPSPR	1147		
Db	1117	PHVQESUSTQRYASADPTVFAPERSPRGELDEEGYWTMRDKPKQEYLNPE-----	1167		
QY	1148	EGPLPAARPAAGATLERAKTLPKNGKGVKVDVFAFGGAVENPEYLTPOGGAAPHPHPPA-	1206		
Db	1168	ENPFVSR-----KNGDLQ-----ALDNPEYHNASNG-----PPKAE	1199		
QY	1207	-----FSPAFDNLVYWDQDPPERGA--PPSTF	1231		
Db	1200	DEYVNEPLYLNTFANTLGAEYLNKNNILSMPEKAKKAFDNPDMYNNHSLPPRSTLQHPDYL	1259		
QY	1232	KGTPT-----AENPEYL	1243		
Db	1260	QESTYKTFYKQNGRIRPIVAENPEYL	1285		
RESULT 8					
S06142					
protein-tyrosine kinase (EC 2.7.1.112) mrk-Y precursor - southern platyfish					
N;Alternate names: epidermal growth factor receptor homolog; kinase-related transforming					
C;Species: Xiphophorus maculatus (southern platyfish)					
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000					
C;Accession: S06142; S13809					
R;Withbrodt, J.; Adam, D.; Malitschek, B.; Maeueller, W.; Raulf, F.; Robertec					
Nature 341, 415-421, 1989					
A;Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu loci					
A;Reference number: S06142; MUID:90015140; PMID:2797166					
A;Accession: S06142					
A;Molecule type: DNA					
A;Residues: 1-1166 <WIT>					
A;Cross-references: EMBL:X16891; NID:g65290; PIDN:CAA34770.1; PID:g65291					
R;Adam, D.; Maeueller, W.; Scharf, M.					
Oncogene 6, 73-80, 1991					
A;Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xiphophorus					
A;Reference number: S13807; MUID:91125882; PMID:1846957					
A;Accession: S13809					
A;Status: preliminary; translation not shown					
A;Molecule type: DNA					
A;Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>					
A;Cross-references: EMBL:X56319; NID:g65284; PIDN:CAA39763.1; PID:g65285					
C;Genetics:					
A;Gene: mrk					
A;Map position: Y					
A;Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1					
C;Superfamily: epidermal growth factor receptor; protein kinase homology					
C;Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein; tyros					
F;1-25/Domain: signal sequence #status predicted <SIG>					
F;26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>					
F;707-972/Domain: protein kinase homology <KIN>					
F;715-723/Region: protein kinase ATP-binding motif					
Query Match					
Best Local Similarity 43.9%; Pred. No. 7.3e-103;					
Matches 557; Conservative 166; Mismatches 399; Indels 146; Gaps 31;					
QY	4	AALCRWGLLLALLPPGAAT----	QVCTGTDMMKRLRASPETHDMLRHLHYQGCVQVQGN	59	
Db	8	AALQ--LLLVLSISCCSTDPDRKVCQGTSNQMTM---LDNHYLKKKKMYSCNVLEN	62		
QY	60	LELTPLTNASISFLQDIOEVGYVLIHNOVROVQLORLIVRGTOLPEDNYVALAVLND	119		
Db	63	LEITYTQENQDLSFLOSIQEVGYVLIHNOVROVQLORLIVRGTOLPEDNYVALAVLND	122		
QY	120	GDPLNNTTFTVGTASPGGLRELQRLSITELKGGVLIQIORNPOLCYQDTILWKDIFHKNQL	179		

Db 123 YQK-NPSSP--DYVQVLKQLQLSNLTLTEILSGGVKSHNPELLCNVETINWMDIVDKTSNP 179
QY 180 ALTLIDNRSRACHPCSPMCKSRGWGSESSDCQSLTRITVCAGCG-ARCKGPLPTDCCHE 238
Db 180 TMLLIHAFERQCKQKDHGCVNGSWAPGPGHCKQKFTKLLCAEQCNRRCRGPKPIDCCNE 239
QY 239 QCAAGCTGPKHSDCLACHHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTA 298
Db 240 HCAGGCTGPRATDCLACRFDNDDGTCCKTPPKIYDIVSHQVVDNPNIKYTFGAACVKE 299
QY 299 CPVNYLSTVGSTLVCPLHNOQVTAEDGTORCEKSKPCARVCYGLGMOYIKANSKF 358
Db 300 CPNRYVUTE-GACVRSAGMLEVD-ENGRSKPCDGVCPKVDGIGISL-SNTIAYN 356
QY 359 ITL-BFAGCKKIFGSLAFELPESFDGDPASNTAPLOPEQLQVFPETLEEITGYLIYISAWPD 417
Db 357 STNIRSFNCTKINGDIILNRNSFEGDPHYKIGITMDPEHLWNLTTVKEITGYLIVIMWPE 416
QY 418 SLPLSVFQNLQVIRGRILHNGAYS-LTQGLGISWGLSLRELGSGLAIHHNTHLCF 476
Db 417 NMTSLSVFQNLQVIRGRITTSRGSFVYVQVRHLQWGLSLKEVSAGNVILKNTLQLRY 476
QY 477 VHTVPWDLFRNPHOALLHANPEDECVCEGLACHOLCARGHCWGPGTQCVCNCSQFLR 536
Db 477 ANTIWRLRFSDDQSIYDART-----ENQTCNNECEDGCV-PGPTMVCSLHVD 528
QY 537 GQCEVBEVRVLOGLPREYNARHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPFC 596
Db 529 GGRCVASCNLLQGEPREAQVDGRCVQCHQECVQTSLTCTCYGPGPANCSSAHFQDGPQC 598
QY 597 VARPCEGVRPDLSYMPIWPKFPPDEGACQPCPINCTHSCVDLDDKGPAPORASPLTSIVS 656
Db 589 IPRCPHGILGDGTL-IWKYADRMGQCQPCQHCNCTQCGSGPLSGCRGD-IVSHSLAVG 646
QY 657 AVVGILLVVLGVVFGILLKRRQKRYKTMRLLOETELVEPLTSGAMPNQAQRILK 716
Db 647 LVSGLLITVIALVILLVLLRRRIK-RKRTIRCLQELQELVELPITSGQAPNFAIRILK 705
QY 717 ETELKRVKVLGSGAFQYIK-----ANSKFIGITELVIRENTSPRANKELIDEAVMAG 770
Db 706 ETEFKDRVLGSGAFGVYKGLWNPGENIRIPVAIKVLREATSPKVNQEVLEDAVMAS 765
QY 771 VGSPPYVSRLLIGICLTSTVQLVTQMPYGCLLDHHVRENRRGLSGQDLLNWCWQIAKMSYL 830
Db 766 VDPHPVCRLLIGICLTSAVQLVTQMPYGCLLDHYRQHOERICQGLLNNWCVQIAKGNYL 825
QY 831 EDVRLVHRDLAARNVLKSNHVKITDFGLARLLDIDETEHADGGKVPKKNWALSILR 890
Db 826 EERHLVHRDLAARNVLLKNPNHVKITDFGLSKULLTADKEKEYQADGGKVPKKNWALSILQ 895
QY 891 RRETHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPPICTIDVYIMV 950
Db 886 WTYTHQSDVMSYGVTVWELMTFGSKYDGIIPAKEIASVLNGLERLPPOPPICTIEVMIL 945
QY 951 KCMWIDSECRPRPRELVSEFSRMARDQPRVVTIONEDLGPASPLDSTFYRSLLEDDMDGD 1010
Db 946 KCMWIDPSSRPRPRELVGSPQARDPSRYLVIQG---NLPSLSDRRLFSLSSD--D 1000
QY 1011 LVDAEBEVLVPOQGFCDPAPGAGGVVHHRSSSTRSGGDLTLGLPSEEAAPRSLA 1070
Db 1001 VVDAEVLVLPYKRI-----NRQGS-----EPCI 1023
QY 1071 PSEGAGSDVFDGDLGMAKGLQSLTHDPSPLOQRYSEDPTV-PLPSETDGYVAPLTCSP 1129
Db 1024 PPTGH-----PVRENSITLRNISDPTQNALEKLDLGH----- 1055
QY 1130 QPEYVNPQVVRPOP-----PSPRE-----GPLP-AARPAGATLERAKTLPSPKNGVVK 1176
Db 1056 --EYVNPQSETSRSLSDIYNPNVEDLTDGWPVLSQSAETNFSRPEYINTNQSL-- 1111
QY 1177 DVFAFGAVENPEYLPQGGAAQPPHPPAFSFAFNLYYWDQDPPERGAPPTFKGTPT 1236
Db 1112 -PLVSSGSMDDPDY---QAG-----YQAAF-----LPQTGALTGNMFLPA 1148

QY 1237 AENPEYL 1244
Db 1149 AENLEYL 1156

RESULT 9

A36223

kinase-related transforming protein (erbb3) (EC 2.7.1.-) precursor - human

A;Species: Homo sapiens (man)

C;Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000

C;Accession: A36223; 159164

R;Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.

Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989

A;Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal g

A;Reference number: A36223; MUID:90083234; PMID:2687875

A;Accession: A36223

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1342 <KRA>

A;Cross-references: GB:M29366

R;Plozman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, G.J

Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990

A;Title: Molecular cloning and expression of another epidermal growth factor receptor-rel

A;Reference number: 159164; MUID:90311312; PMID:2164210

A;Accession: 159164

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>

A;Cross-references: GB:M34309; NID:G183990; PIDN:AAA35979.1; PID:G306841

C;Genetics:

A;Gene: GDB:ERBB3; HER3

A;Cross-references: GDB:119880; OMIM:190151

A;Map position: 12q13-12q13

C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog

C;Keywords: ATP; phosphotransferase

F;707-972/Domain: protein kinase homology <KIN>

F;715-723/Region: protein kinase ATP-binding motif

Query Match 34.0%; Score 2305.5; DB 2; Length 1342;
Best Local Similarity 39.5%; Pred. No. 3.9e-92;
Matches 519; Conservative 191; Mismatches 463; Indels 141; Gaps 35;

QY 10 GLLLALLPPGAA--STQVCTGTDMLKRLPASPEHLMLRLHYQGCVVQGNLELTYLPT 67

Db 11 GLLFSLARGSEVGNVSOAVCPGTGLNGLSVTGDAENQYQTLYKLYERCEVVMGNLEIVLTGH 70

QY 68 NASLSFLQDIQVQGVVLIHAHQVQVPLQRLIRVGTQOLFEDNYALAVLDNGDPLNNTT 127

Db 71 NADLSFLOWIREVTGYVLVAMNEFSLPLNLRVVRGTQVYDGKFAIFVM-----LYNNT 125

QY 128 PVTGASPGGLREQLRSLTEILKGGVLIQORNPOLCVQDTILMKDIFHKNQLALTLDTN 187

Db 126 ----NSSHALRQLRLTQLTEILSGGVYIEKNOKLCHMDITDWRDIVVRD--AEIVVKD 178

QY 188 RSRACHPCSPMCKSRGWGSESSDCQSLTRITVCAGCG-ARCKGPLPTDCCHECAAGCTG 246

Db 179 NGRSCPPCHEVCKG-RCWGPGSEDCQTLTKTICAPQCNHCGFGPNQCCHECAGCSG 237

QY 247 PKHSDCLACHNHSIGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACPYNYLST 306

Db 238 PQDITDCFACRHFENDSGACVPRCPQPLVYNKLTFOLEPNPHTKYQYGGVCVASCAPHNFV-V 296

QY 307 DVGSCCTLVCLPHNQEVTAEDGTORCEKSKPCARVCYGLGMOYIKANSKF--IGITELE- 363

Db 297 DOTSCVRACPPPKMEVD-KNGLKMCPECCGLCPKACEGTG-----SGSRPQTVDSSNIDG 350

QY 364 FAGCKKIFGSLAFELPESFDGDPASNTAPLQPBQLQVFTLEEITGYLIYISAWPDSLPDLS 423

Db 351 FVNCTKILGNLDFLLTGLNGDPWHKIPALDPEKLVFRVREITGYLNTQSPPHMHNFS 410

QY 424 VFQNLQVIRGRILHNGAYS-LTQGLGISWGLSLRELGSGLAIHHNTHLCFVHTVPW 482

Db 411 VFSNLTIGRSLSYNRGFSLLIMKNLVNTSLGFRSLKEISAGRIYISANRQLCVHHSLNW 470
QY 483 DQFRPNQOALLHTA-NRPEDECVGEGLACHOLCARGHGWGPGTQCVCNCSQFLRGQECV 541
Db 471 TKVLRGTEERLIDIKHNRPRDCVABEKVCDPLCSSGCGWPGPGQCLSCRNVSRGVCV 530
QY 542 EECRVLOGLPREYVYVNRHCLPCHPECPQNGSVTFCFGEADQCACAHYKDPDFCVARCP 601
Db 531 THCNFLNGEPREFAHEAECSCHPECPQMGEGTATCNGSGSDTCAQCAHFRDGHPCVSSCP 590
QY 602 SGVKPDLSPYMIKFPDEBEGACQCPINCHTHSCVDLDDKCPABQRA-----SPLTSIVSA 657
Db 591 HGVLG--AKGPIYKYPDVQNECRCHENCCTGCKGPELODCLGQTLVLKIKTHLTWALT 648
QY 658 VVGILLVVLGVVFGILLIKRQOKIR-KYTMRRLLQETELVEPLTPSGAMPNQAMRILK 716
Db 649 IAG--LVVIFMVLGGTFLYWRGRQIONKRAMRRYLERGESIEPLDPS-EKANKVLARIFK 705
QY 717 ETELKVKVLGSGAFQYIK-----ANSKFIGITELVLRNTSPKANKELDEAYVMAG 770
Db 706 ETELKVLKVLGSGVFGTVHKGWIPGESIKIPVCIKVIEDKSGRQSFQAVTDHMLAIGS 765
QY 771 VGSFYVSRLLIGICLTSTVOLTPQLMPYGCLLDHYRNRGRGLGSDLLNWCMTAKGMSYL 830
Db 766 LDHAHIVRLGLCPGSSQLVTOYPLPLGSLDHYRQHRGALGQLLLNMGVQIAKGMYYL 825
QY 831 EDVRLVHRLDAARNVLKSPNHVKITDFGLARLLDIDETVHADGGKVPKMMALESILR 890
Db 826 EHGWHNRNLAARNVLKSPQOVADFGVADLLPPDDKQLLYSEAKTPKMALESIH 885
QY 891 RRFTHQSDVMSYGVYVWELMTFGAPYDGIPIAREIPDLLEKGERLPQPICTIDVYIMV 950
Db 886 GKYTHQSDVMSYGVYVWELMTFGAEPVAGLRLAEVPLLEKGERLAQPICTIDVYIMV 945
QY 951 KCMWIDSECPREFELVSFRWARDQRFVVIQNEDLGPA---SPLDSTFYRSLLEDD 1007
Db 946 KCMWIDENIRPTEKELANEFTRWARDPPRYLVIKRES-GPGIAPGPEPHGLTNKCLEVE 1004
QY 1008 MGDVDAEYLVPOQGFCDPAPGAGGMVHRRSSSTRSGGDLTLGLEP-SEBEAPR 1066
Db 1005 LEPELDLDLEAED-----NLATTLGSLALSPVGTLLNRRPGSQ 1045
QY 1067 SPLAPSEGASDVFDGLMGAAKGLQSLPHTD-PSPLQRYSEDPTVPLP-----SETD 1119
Db 1046 SLLSPSSGY-MPMNQNGLGSCQESAVSGSSRCPFPVSLH-----PMRPGCLASESE 1098
QY 1120 GYVA-----PLTCSPOPE-----YVNPQDVVRPQPPSPREP----- 1150
Db 1099 GHVTGSEAELEQKYSMCRSRSRSPRPGDSAYHSQRHSLTPVTPPLSPGLEEEDVNG 1158
QY 1151 --LPAARPAGATLERAKTSLP-GKNGVV-----KQVFAFGGAVENPEYLTPOGGAAPQ 1201
Db 1159 YVNPDTHLKTPSREGTLLSVGLUSSVLGTEEBED-----EYEVNRRRRRSP-P 1209
QY 1202 HPPAPFAPFDNLYYND-----QDPERGAPPSTFKGTPTAENPEYL 1243
Db 1210 HPPRPSLEELGVEYMDVGSLSASLGTSQCLPLHPVPIINPTAGTTPDEDEYEM 1263

RESULT 10
JC4387
epidermal growth factor receptor homolog precursor - rat
N;Alternate names: ErbB3 protein; HER3 protein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C;Accession: JC4387
R;Hellyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A;Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.
A;Reference number: JC4387; MUID:96036535; PMID:8522190
A;Accession: JC4387
A;Molecule type: mRNA
A;Residues: 1-1339 <HEL>

A;Cross-references: GB:U29339; NID:G915389; PID:G915390
A;Experimental source: liver
C;Note: The authors translated the codon AAC for residue 369 as Thr and GTT for residue 370
C;Comment: This protein is a functional heregulin receptor that transduces signals to the nucleus
A;Gene: ErbB3
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
C;Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1339/Product: epidermal growth factor homolog #status predicted <TM>
F;640-659/Domain: transmembrane #status predicted <TM>
F;705-970/Domain: protein kinase homolog <KIN>
F;713-721/Region: protein kinase ATP-binding motif
F;939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr) (cat)
Query Match 32.7%; Score 2218.5; DB 2; Length 1339;
Best Local Similarity 39.8%; Pred. No. 2.1e-88;
Matches 509; Conservative 168; Mismatches 443; Indels 165; Gaps 36;
QY 3 LAALCRWGLLLALLPPGAA---STVCTGTDMKRLPASPETHLDMRLHLYQCQCVVQGN 59
Db 7 LQVLC---FLLSLARGSEMGNSQAVCPGTNLGLSVTGADNQYQTYLYKLYKECEVMGN 62
QY 60 LEITYLPTNASLSFLQDIOEVQGYVLIHNRQVQLRILRIVRGTLQFEDNVALAVLDN 119
Db 63 LEIVLTGHNADLSFLQWIREVTAYVLVAMNEFSVLPPLNLRVVRGTVQVYDGKEAIFVM-- 120
QY 120 GDLPLNNTTPTVGTASPGGLRELQRLSITELKGGVLTORNPOLCYQDTILWKDIFHKNNQL 179
Db 121 ---LNTYNT---NSSHALRQLKFTQLTEILSGGVYIEKNDKLMCHMDTIDHRDLVRVR--- 170
QY 180 ALTLIDNRSRACHPCSPMKGSRGSESSDCQSLTRTVCAAGC-ARCKGPLPTDCHE 238
Db 171 GAEIVKNGANGANGCPPCHEVCKG-RCWGPGPDCCQLITKTCAPQCNCRGFGPNQCCHD 229
QY 239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTTFESMPNPEGRYTFGASCVTA 298
Db 230 ECAGGCGSQDTCDFACRRFNDSGACVPRCPPLVYNKLTFOLEPNPHTKYQYGGVCVAS 289
QY 299 CPYNYLSTDVGCTLVCPHNLQEVTAEDGTORCEKSKPCARVCYL--GMQYIKANSKF 356
Db 290 CPNHFV-VDTPTFCVRACPPDPKMEVD-KHGLKMCPECGGLCPKACEGTSGSRYTQYVDSN 347
QY 357 IGITELEFAGCKKIFGSLAPLPESPDGPDASNTAPLOEQLOVFTLEETITGLYISAWP 416
Db 348 ID----GFVNCYKILGNLDFLITGLNVDPAWKIPALDPEKLNVRVIRVITGLYINTQSWP 403
QY 417 DSLPDLSPFQNLQVIRGRILHNGAYS-LTQGLGISWGLRSLRSLRELGLALHNNHLC 475
Db 404 PHMHNFVSFNLTTIGRSLYNRGFSLLIMKNLVNTSLGFRSLKEISAGRIYISANRQLC 463
QY 476 FVHTVPWDOLFNRPHQALLHTA-NRPEDECVGEGLACHOLCARGHGWGPGTQCVCNCSQF 534
Db 464 YHSLNWTLLRGLPSEERLDIKYDRPLGECLAEGKYCDPLCSSGCGWPGPGQCLSCRNY 523
QY 535 LRQECVECEVRVLQGLPREYVYVNRHCLPCHPECPQNGSVTFCFGEADQCACAHYKDPDP 594
Db 524 SREGCVTHCNFLOGEPREFVHEAQCFSCHPCECLPMEGTSTYNGSGSDACARCAHFRDGP 593
QY 595 FCVARCPGSKVCPDLSYMPIWKFPDEBEGACQCPINCHTHSC--VDLDDKCPABQRAASPLT 652
Db 584 HCVNSCPHGILG-AKGPIYKYPDAQNECRPCHENCCTGCKGPELODCLGQAEVLMKSPH 641
QY 653 SIVSAVVGILLVVLGVVFGILLIKRQOKIR-KYTMRRLLQETELVEPLTPSGAMPNQAO 711
Db 642 LVIAVTG--LAVILMLGSSFLYWRGRQIONKRAMRRYLERGESIEPLDPS-EKANKVL 698
QY 712 MRILKETELRKVKVLGSGAFQYIK-----ANSKFIGITELVLRNTSPKANKELDEA 765
Db 699 ARIFKETELRKVLKVLGSGVFGTVHKGWIPGESIKIPVCIKVIEDKSGRQSFQAVTDHM 758
QY 766 YVNVAGVSPYSRLLIGICLTSTVOLTPQLMPYGCLLDHYRNRGRGLGSDLLNWCMTAK 825

Query Match	22.7%	Score 1537;	DB 2;	Length 544;
Best Local Similarity	52.5%;	Pred. No. 2e-59;		
Matches 330;	Conservative	69;	Mismatches 131;	Indels 98;
Gaps	16;			
QY	578	GPEADQVCACAHYKDPFCVCARCPGKVPDLSPYMPKPPDESGACOPCPINCHTSCVDL	637	
Db	1	GP--DHCMKCAHFDGCHPCVACAGVGLGENDTL-VMKYADANAVCOLCHPNCITRGCKGP	57	
QY	638	DDKGPACQASPLTISVSJAV--GILLVVVLGVVFGILIKRROOKIRKYTMRLLOQETEL	696	
Db	58	GLEGCP--NGSKTPSIAGVVGGLLCLVVGLIGLGLYLR--HIVKRTLRLLQEREL	113	
QY	697	VEPLTPSGAMPNOAQRILKETELRKVKVLGSGAFQYIK-----ANSKFIGITELVLR	750	
Db	114	VEPLTPSGEAPNOAHLRIKETEFKKVKVLGFGAFGVYKGLMIPGEKVTIPVAIKELR	173	
QY	751	ENTSPKANKIIDEAYVMAGVSGPYYSRLGICLTSTVQLVLTOLMPYGCLLDHVRENGR	810	
Db	174	EATSPKANKIIDEAYVMASVDMPHVCRLLGICLTSTVQLITQLMPYGCLLDIYREHKON	233	
QY	811	LGSODLLNMCQIAKMSYLEDVLRVLRDLAARNVLVKSPNHVKITDFGLARLLDIDETE	870	
Db	234	IGSQYLLNMCVQIAKGMVYLEERHVRDLAARNVLVKTQHVKIIDFGLAQLGADEKE	293	
QY	871	YHADGGKVPITKMALESILRRRTHOSDWSYGYVTWELMTFGAKPYDGIIPAREIPDLLE	930	
Db	294	YHAEGGKVPITKMALESILHRIYTHOSDWSYGYVTWELMTFGSKPYDGIIPASEISSVLE	353	
QY	931	KGRLPOPPCTIDVYIMVWKWIDSECPREFELVSEFSRMARDPQRFVVIQ--NEDLG	989	
Db	354	KGRLPOPPCTIDVYIMVWKWMSDADSPKREFLIAEFSKWARDPPYVLVQLGDERMH	413	
QY	990	PASPLDSTFYRSLLEDDMGDLVDABEYLVPQQGFCCPDPAAGAGGVHRRHSSSTRSG	1049	
Db	414	LPSPDTSKPYRTLMEEDMEDIVDABEYLVPHQGFF-----NSPST---454		
QY	1050	GGDLTLGLEPSEEARPSL-----APSEGAGSDVFGDGLGMGAACKLQSLPHTHDSPLQ	1104	
Db	455	-----SKTPLLSSLATSNNSATNCIDRNGG-----H-----481		
QY	1105	RYSEDPVPLPSETDGTGVAPLTCSPQEVYVQDVRPQPPSPREGPLPAARPAAGAT--LER	1163	
Db	482	-----PVREDGFL-----PAPEYVQ--LMPKKPSTAMVQNIYVYLSLTAISK	523	
QY	1164	AKTLPSPKGVKVDVFAFGGAVENPEYL	1191	
Db	524	LPIDSRYN-----SHSTAVDNPEYL	544	

RESULT 14

S00727
 kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis virus
 C:Species: avian erythroblastosis virus
 C:Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
 C:Accession: S00727
 R:Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
 Oncogene Res. 1, 265-278, 1987
 A:Title: Common site of mutation in the erbB gene of avian erythroblastosis virus
 A:Reference number: S00727; MUID:88217326; PMID:2897102
 A:Accession: S00727
 A:Molecule type: DNA
 A:Residues: 1-545 <SCO>
 A:Cross-references: EMBL:X06943
 C:Genetics:
 A:Gene: erbB
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: Anp; phosphotransferase
 F:135-400/Domain: protein kinase homology <XIN>
 F:143-151/Region: protein kinase ATP-binding motif
 Query Match 22.6%; Score 1530; DB 2; Length 545;
 Best Local Similarity 52.5%; Pred. No. 3, 9e-59;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:47:04 ; Search time 10.5911 Seconds
(without alignments)

4891.279 Million cell updates/sec

Title: SEQ4-710-730-12

Perfect score: 6775

Sequence: 1 MELALCRWGLLLALLPPGA.....TFKGTPAENPEYGLDVPV 1249

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	6602	97.4	1255	1 ERB2 HUMAN	P04626 homo sapien
2	5811	85.8	1257	1 ERB2 RAT	P06494 rattus norv
3	5795.5	85.5	1254	1 ERB2 MESAU	P00553 mesocricetu
4	3028	44.7	1210	1 EGFR HUMAN	P00533 homo sapien
5	3002	44.3	1210	1 EGFR MOUSE	Q01279 mus musculu
6	2849.5	42.1	1308	1 ERB4 HUMAN	Q15303 homo sapien
7	2830	41.8	1308	1 ERB4 RAT	Q62956 rattus norv
8	2569.5	37.9	1167	1 XMRK_XIPMA	P13388 xiphophorus
9	2314.5	34.2	1342	1 ERB3 HUMAN	P21860 homo sapien
10	2241.5	33.1	1339	1 ERB3 RAT	P062799 rattus norv
11	1845	27.2	1426	1 EGFR DROME	P04412 drosophila
12	1633.5	24.1	634	1 ERBB ALV	P00534 avian leuko
13	1595	23.5	703	1 EGFR CHICK	P13387 gallus gall
14	1588	23.4	604	1 ERBB AVIER	P00535 avian eryth
15	1520	22.4	540	1 ERBB AVIEU	P11273 avian eryth
16	1224	18.1	1323	1 LTB3 CAEEL	P24348 caenorhabdi
17	1142.5	16.9	245	1 ERB2 MOUSE	P70424 mus musculu
18	703	10.4	1363	1 ILPR BRALA	Q02466 branchiost
19	676.5	10.0	1372	1 INSR_MOUSE	P15208 mus musculu
20	674	9.9	1383	1 INSR RAT	P15127 rattus norv
21	673	9.9	1382	1 INSR HUMAN	P06213 homo sapien
22	659	9.7	1300	1 IRR_MOUSE	Q9wt14 mus musculu
23	650	9.6	1297	1 IRR_HUMAN	P14616 homo sapien
24	644.5	9.5	1300	1 IRR_CAVPO	P14617 cavia porce
25	643	9.5	1477	1 HTK7_HYDAT	Q25197 hydra atten
26	634	9.4	1607	1 MPR_LYNST	Q25410 lymanaea sta
27	607	9.0	1367	1 IG1R HUMAN	P08069 homo sapien
28	589	8.7	1373	1 IG1R_MOUSE	Q60751 mus musculu
29	586	8.6	987	1 EPB4_HUMAN	P54760 homo sapien
30	585.5	8.6	1370	1 IG1R RAT	P24062 rattus norv
31	583	8.6	1114	1 RET_HUMAN	P07949 homo sapien
32	574.5	8.5	984	1 EPB1_CHICK	Q07494 gallus gall
33	573	8.5	2146	1 INSR_DROME	P09208 drosophila

RESULT 1	ERB2_HUMAN	STANDARD;	PRT;	1255 AA.
AC	P04626;			
DT	13-AUG-1987 (Rel. 05, Created)			
DT	13-AUG-1987 (Rel. 05, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)			
DE	(p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell			
DE	surface receptor HER2) (MLN 19).			
GN	ERBB2 OR HER2 OR NGL OR NEU.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86118663; PubMed=3003577;			
RA	Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,			
RA	Saito T., Toyoshima K.;			
RT	"Similarity of protein encoded by the human c-erb-B-2 gene to			
RT	epidermal growth factor receptor.";			
RL	Nature 319:230-234(1986).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86070181; PubMed=2999974;			
RA	Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,			
RA	McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,			
RA	Frankie U., Levinson A., Ullrich A.;			
RT	"Tyrosine kinase receptor with extensive homology to EGF receptor			
RT	shares chromosomal location with neu oncogene.";			
RL	Science 230:1132-1139(1985).			
RN	[3]			
RP	SEQUENCE OF 737-1031 FROM N.A.			
RX	MEDLINE=86016729; PubMed=2995967;			
RA	Semba K., Kamata N., Toyoshima K., Yamamoto T.;			
RT	"A v-erbB-related protooncogene, c-erbB-2, is distinct from the			
RT	c-erbB-1/epidermal growth factor-receptor gene and is amplified in a			
RT	human salivary gland adenocarcinoma.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).			
RN	[4]			
RP	VARIANTS VAL-654 AND VAL-655.			
RX	MEDLINE=93194196; PubMed=8095488;			
RA	Ehsani A., Low J., Wallace R.B., Wu A.M.;			
RT	"Characterization of a new allele of the human ERBB2 gene by allele-			
RT	specific competition hybridization.";			
RL	Genomics 15:426-429(1993).			
CC	!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,			
CC	ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A			
CC	POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-			
CC	ALPHA AND AMPHIREGULIN.			
CC	!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein			
CC	tyrosine phosphate.			
CC	!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS			
CC	(POTENTIAL).			
CC	!- SUBCELLULAR LOCATION: Type I membrane protein.			

P09759 rattus norv
Q93105 aedes aegypt
P54761 mus musculu
P54762 homo sapien
Q91736 xenopus lae
Q91571 xenopus lae
Q03145 mus musculu
P29317 homo sapien
Q91694 xenopus lae
P54753 homo sapien
Q91845 xenopus lae
P18460 gallus gall

ALIGNMENTS

FT DISULFID 569 BY SIMILARITY.
FT DISULFID 589 BY SIMILARITY.
FT DISULFID 605 BY SIMILARITY.
FT DISULFID 625 BY SIMILARITY.
FT DISULFID 636 BY SIMILARITY.
FT DISULFID 644 BY SIMILARITY.
FT MOD RES 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD RES 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 188 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 260 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 631 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;
Query Match 85.8%; Score 5811; DB 1; Length 1257;
Best Local Similarity 85.6%; Pred. No. 4.7e-304;
Matches 1077; Conservative 53; Mismatches 118; Indels 10; Gaps 4;
QY 1 MELAALCRWGLLALLPPGAASCTGCTDMKRLPASPETHLDMLRHLVGGCQVVOGNL 60
DB 1 MELAALCRWGLLALLPPGAASCTGCTDMKRLPASPETHLDMLRHLVGGCQVVOGNL 60
QY 61 ELTYLPTNASLFLQDIQEVGYVLIHANOVRQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
DB 61 ELTYVPAASLFLQDIQEVGYVLIHANOVRQVPLQRLRIVRGTLQFEDKYALAVLDNR 120
QY 121 DPLNNTPTVT-GASPGGLRELQRLSLTEILKGGVLIQRLNOLCYQDTILKWDIPHKNQOL 179
DB 121 DPQDNVAASPTGRTPEGLRELQRLSLTEILKGGVLIQRLNOLCYQDMVLWKVDVPRKNQOL 180
QY 180 ALTLIDNRSRACHPCSPMCKGSCWGESSEDCOSLRTVTCAGGCARCKGLPTDDCCHEQ 239
DB 181 APVDIDNRSRACHPCSPMCKGSCWGESSEDCOSLRTVTCAGGCARCKGLPTDDCCHEQ 240
QY 240 CAAGCTGPKHSDDLACLFHNSHGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
DB 241 CAAGCTGPKHSDDLACLFHNSHGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTTC 300
QY 300 PYNVLTSDVSGCTLVCPHNOEVTAEQGTCEKSKPCARVCYGLGMOYIKANSKFIGI 359
DB 301 PYNVLTSEVSGCTLVCPHNOEVTAEQGTCEKSKPCARVCYGLGMEHLRGARATSD 360
QY 360 TELEFACGKIFGSLAFIPESFDGDRASNTAPLOPEOLQVPETLEETGVLYISAWPDSL 419
DB 361 NVQFEDGCKIFGSLAFIPESFDGDRASNTAPLOPEOLQVPETLEETGVLYISAWPDSL 420
QY 420 PDLVSFQNLQVIRGILHNGAYSILTLQGLGISWLGRLSLRELGLSLALIHNTHLFCVHT 479
DB 421 RDLVSFQNLRIIRGILHNGAYSILTLQGLGHSGLSLRELGLSLALIHNTHLFCVHT 480
QY 480 VPWDLFRNPHQALLHTANRPEDE-CVGEGLACHQLCARHCWGPGPTQCVCNCSQFIRGQ 538
DB 481 VPWDLFRNPHQALLHTANRPEDE-CVGEGLACHQLCARHCWGPGPTQCVCNCSHFIRGQ 540
QY 539 ECVEECRVLCGLPREYNARHCLCHPECPQNGSVTCFGEADQCVACAHYKDPPECVA 598
DB 541 ECVEECRVLCGLPREYNARHCLCHPECPQNGSVTCFGEADQCVACAHYKDPPECVA 600
QY 599 RCPGSGVPLDSYMPIWFFPDEGACQPCINCHTSCVDLDDKGCAPABQASPLTISVAV 658
DB 601 RCPGSGVPLDSYMPIWFFPDEGACQPCINCHTSCVDLDDKGCAPABQASPLTISVAV 660
QY 659 VGILLVVVLGVGILIKRQOKIRKTYMRLLQETELVEPLTSPGAMPQAOQRILKET 718
DB 661 VGILLVVVLGVGILIKRQOKIRKTYMRLLQETELVEPLTSPGAMPQAOQRILKET 720
QY 719 ELRKVKVLGSAFGQYKA-----NSKFTIGITELVIRENTSPKANKEILDEYVAVG 771
DB 721 ELRKVKVLGSAFGTVYGIWIPDGENVK-IPVAIKVIRENTSPKANKEILDEYVAVG 779
QY 772 GSPYVSRLLGICLTSTVQLVTPMYPYGLLDHVRGRGLSGQDLLNWCQIAKGMYSLE 831

DB 780 GSPYVSRLLGICLTSTVQLVTPMYPYGLLDHVRGRGLSGQDLLNWCQIAKGMYSLE 839
QY 832 DVELVHRDLAARNVLKSPNHVKITDPLGLARLLDIDETEHADGGKVPKMMALESILRR 891
DB 840 DVELVHRDLAARNVLKSPNHVKITDPLGLARLLDIDETEHADGGKVPKMMALESILRR 899
QY 892 RFTHQSDVMSYGVYVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYIMVK 951
DB 900 RFTHQSDVMSYGVYVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYIMVK 959
QY 952 CWMIDSECRPRRELVSERFMRDQRFVVIQNEIDLPASPLDSTFYSLLEDGMDGL 1011
DB 960 CWMIDSECRPRRELVSERFMRDQRFVVIQNEIDLPASPLDSTFYSLLEDGMDGL 1019
QY 1012 VDAREEYLVPOQGFPCPDPAAGMVGHHRRSSSTRSGGDLTLGLEPSEEAAPRSLAP 1071
DB 1020 VDAREEYLVPOQGFPCPDPAAGMVGHHRRSSSTRSGGDLTLGLEPSEEAAPRSLAP 1079
QY 1072 SEGAGSDVFDGDLGMAAGKQLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQP 1131
DB 1080 SEGAGSDVFDGDLGMAAGKQLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQP 1139
QY 1132 EYVNPQDVPVPPSPREGPLPAARPAAGATLERAKTLSPKNGVVKOVFAFGAVENPEYL 1191
DB 1140 EYVNPQDVPVPPSPREGPLPAARPAAGATLERAKTLSPKNGVVKOVFAFGAVENPEYL 1199
QY 1192 TPQGGHAPQHPHPPAFSPADNLYYWDQPPPERGAPSTFKGTPTAENPEYLGUDVVP 1249
DB 1200 VPREGTASPHPSPAFSPADNLYYWDQPPPERGAPSTFKGTPTAENPEYLGUDVVP 1257
RESULT 3
ERB2_MESAU STANDARD; PRT; 1254 AA.
ID ERB2_MESAU Q60553;
AC Q60553;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (BC 2.7.1.112)
DE (P1888B2) (NEU proto-oncogene) (C-erbB-2).
GN ERB2 OR NEU.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RX MEDLINE=94153007; PubMed=7908275;
RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255 (1994).
CC - FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC - SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC - SUBCELLULAR LOCATION: Type I membrane protein.
CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC	EMBL; D16295; BAA03801.1; -	QY	61	ELTYLPTNASLSFLQDIOEVQGYVLIHNRQVPLQRLRIVRGTLQFEDNYALAVLDNG	120
CC	HSP; P11362; IFGK.	DB	61	ELTYLPANATLSFLQDIOEVQGYVLIHNRQVPLQRLRIVRGTLQFEDNYALAVLDNR	120
DR	InterPro; IPR000494; EGFR_L domain.	QY	121	DPLNNTPTVGTASPGGLRELRSLTEILKGGVLIORNPQLCYQDTILWKDIFKHKNOLA	180
DR	InterPro; IPR000719; Euk_pkinase.	DB	121	DPLDNVTTATGTPEGLRELRSLTEILKGGVLIORNPQLCYQDTILWKDIFKHKNOLA	180
DR	InterPro; IPR001245; Tyr_pkinase.	QY	181	LTLIDNRSRACHPCSPMCKGSRWCSESDCQSLRTRVCAGGCARCKGLPTDCCHEQC	240
DR	Pfam; PF00069; pkinase; 1.	DB	181	PVDIDNRSRACHPCSPMCKGSRWCSESDCQSLRTRVCAGGCARCKGLPTDCCHEQC	240
DR	Pfam; PF00757; Furin-like; 1.	QY	241	AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMENPEGRYTFGASCVTAC	300
DR	Pfam; PF01030; Recep_L domain; 2.	DB	241	AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMENPEGRYTFGASCVTTC	300
DR	SMART; SM00261; FU; 3.	QY	301	YNYLSTDVGSCITLVCPHNOEVTAEADGTORCEKSCPCARVCYGLGMQVYKANSKFIGIT	360
DR	SMART; SM00219; Tyrc; 1.	DB	301	YNYLSTDVGSCITLVCPHNOEVTAEADGTORCEKSCPCARVCYGLGMQVYKANSKFIGIT	360
DR	PROSITE; PS00107; PROTEIN KINASE_ATP; 1.	QY	361	ELEFACKKIFGSLAFLPESFDGDPASNTAPLOPELOQVFETLEETITGLYIISAWPDSL	420
DR	PROSITE; PS00109; PROTEIN KINASE_TYR; 1.	DB	361	IQEFACKKIFGSLAFLPESFDGDPASNTAPLOPELOQVFETLEETITGLYIISAWPDSL	420
DR	PROSITE; PS00111; PROTEIN KINASE_DOM; 1.	QY	421	DLVSFQNLQVIRGRILHNGAYSILTLQGLISWLGSLRSLRELGSGLALIHHTHLCFVHTV	480
DR	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	DB	421	DLVSFQNLQVIRGRILHNGAYSILTLQGLISWLGSLRSLRELGSGLALIHHTHLCFVHTV	480
DR	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	QY	481	PMDQFRNPHQALLHTANRPECEVGEGLACHOLCARGHCWGPGTQCVNCSOFLRGQEC	540
DR	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	DB	481	PMDQFRNPHQALLHTANRPECEVGEGLACHOLCARGHCWGPGTQCVNCSOFLRGQEC	540
DR	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	QY	541	VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFCVARC	600
DR	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	DB	541	VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFCVARC	600
DR	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	QY	601	PSGVKPDLSYMPIWKYPDEEGMCQPCINCHTSCVDLDDKCPAEQASPLTSIVSAVVG	660
DR	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	DB	601	PSGVKPDLSYMPIWKYPDEEGMCQPCINCHTSCVDLDDKCPAEQASPLTSIVSAVVG	660
DR	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	QY	661	ILLVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILKETEL	720
DR	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	DB	661	ILLVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILKETEL	720
DR	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	QY	721	RKVKVLGSGARGQYKA-----NSKFGITELVRENTSPKANKEILDEAYVMAGVGS	773
DR	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	DB	721	RKVKVLGSGARGQYKA-----NSKFGITELVRENTSPKANKEILDEAYVMAGVGS	773
DR	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	QY	774	PYVSRLGLCLTSTVOLVTLMPYGCCLLDHVRNRLGSLQDILLNWCMTAKGMSYLEDV	833
DR	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	DB	774	PYVSRLGLCLTSTVOLVTLMPYGCCLLDHVRNRLGSLQDILLNWCMTAKGMSYLEDV	833
DR	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	QY	834	RLVHRLDAAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPFKWMALESILRRRF	893
DR	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	DB	834	RLVHRLDAAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPFKWMALESILRRRF	893
DR	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	QY	894	THQSDVMSYGVVWELMTGAKPYDGIIPAREIPDLLEKGERLPPOPICITDVMYMKWC	953
DR	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	DB	894	THQSDVMSYGVVWELMTGAKPYDGIIPAREIPDLLEKGERLPPOPICITDVMYMKWC	953
DR	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	QY	954	MIDSECRPRFRLVSEFSRWARDPQRFVVIQNEIDLGSPASPLDSTFYRSLLEDDMDGLVD	1013
DR	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	DB	954	MIDSECRPRFRLVSEFSRWARDPQRFVVIQNEIDLGSPASPLDSTFYRSLLEDDMDGLVD	1013
DR	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	QY	1014	ABEYLVPQGGFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEBEAPRSLAPSE	1073
DR	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	DB	1014	ABEYLVPQGGFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEBEAPRSLAPSE	1073
DR	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	QY	1074	GAGSDVFDGDLGMAAGKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEY	1133
DR	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	DB	1074	GAGSDVFDGDLGMAAGKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEY	1133
DR	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	QY	1080	GAGSDVFDGDLGMAAGKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEY	1133
DR	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	DB	1080	GAGSDVFDGDLGMAAGKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEY	1133

Query Match 85.5%; Score 5795.5; DB 1; Length 1254;
 Best Local Similarity 85.2%; Pred. No. 3.2e-303;
 Matches 1070; Conservative 60; Mismatches 115; Indels 9; Gaps 3;
 1 MELAAALCRWGLLALLPAGAASTQVCTGTOMKLRASPETHLDMRLHYQGCQVQGNL 60
 1 MELAAWCGWGLLALLPAGAASTQVCTGTOMKLRASPETHLDMRLHYQGCQVQGNL 60

OY 1134 VNQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVYKDVAFGGAVENPEYLTP 1193
 Db 1140 VNQEVPRQPLLTTEGPLPPVRPAAGATLERKTLSPGKNGVYKDVAFGGAVENPEYLVP 1199
 OY 1194 QGGAAPQHPHPAPSPAFDNLVYWDQPPERGAPSPFTKGTPTAENPEYGLDVPV 1249
 Db 1200 RGGASOPH-PPALCPAFDNLVYWDQPPSERGSPFTFGTPTAENPEYGLDVPV 1254
 RESULT 4
 EGFR HUMAN
 ID EGFR HUMAN STANDARD; PRT; 1210 AA.
 AC P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; Q00732;
 AC O00688; Q9B2S2; Q9B2C9; Q9G2X1; Q9H3C9;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
 DE protein-tyrosine kinase ErbB-1).
 GN EGFR OR ERBB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=84219729; PubMed=6328312;
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
 RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
 RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant
 RT expression of the amplified gene in A431 epidermoid carcinoma cells.";
 RL Nature 309:418-425(1984).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 RX MEDLINE=95382957; PubMed=7654368;
 RA Ilekis J.V., Stark B.C., Scoccia B.;
 RT "Possible role of variant RNA transcripts in the regulation of
 RT epidermal growth factor receptor expression in human placenta.";
 RL Mol. Reprod. Dev. 41:149-156(1995).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 RX MEDLINE=97078686; PubMed=8919811;
 RA Reiter J.L., Mahle N.J.;
 RT "A 1.8 kb alternative transcript from the human epidermal growth
 RT factor receptor gene encodes a truncated form of the receptor.";
 RL Nucleic Acids Res. 24:4050-4056(1996).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 RX MEDLINE=97256547; PubMed=9103388;
 RA Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;
 RT "Expression of a truncated epidermal growth factor receptor-like
 RT protein (TEGFR) in ovarian cancer.";
 RL Gynecol. Oncol. 65:36-41(1997).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
 RX TISSUE=Placenta;
 RX MEDLINE=21100872; PubMed=11161793;
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
 RA Schehl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
 RA Balasubramanian S., Crossley T.R., Magnuson T.R., James C.D.,
 RA Mahle N.J.;
 RT "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative EGFR transcripts encoding truncated receptor
 RT isoforms.";
 RL Genomics 71:1-20(2001).
 RN [6]
 RP SEQUENCE OF 575-687 FROM N.A.
 RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schehl C.M.,
 RA Lampland A.L., Balasubramanian S., Crossley T.O., Magnuson T.R.,

RA Mahle N.J.;
 RT "Human and mouse alternative EGFR transcripts encoding only the
 RT extracellular domain of the receptor.";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE OF 713-924 FROM N.A.
 RX MEDLINE=84196372; PubMed=6326261;
 RA Lin C.R., Chen W.S., Krueger W., Stolarsky L.S., Weber W.,
 RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
 RT "Expression cloning of human EGF receptor complementary DNA: gene
 RT amplification and three related messenger RNA products in A431
 RT cells.";
 RL Science 224:843-848(1984).
 RN [8]
 RP SEQUENCE OF 150-962 FROM N.A.
 RX MEDLINE=84245835; PubMed=6330563;
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
 RA Roe B.A., Merlino G.T., Pastan I.;
 RT "Human epidermal growth factor receptor cDNA is homologous to a
 RT variety of RNAs overproduced in A431 carcinoma cells.";
 RL Nature 309:806-810(1984).
 RN [9]
 RP SEQUENCE OF 1028-1210 FROM N.A.
 RX MEDLINE=85046483; PubMed=6093780;
 RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
 RA O'Malley B.W.;
 RT "Isolation of an evolutionarily conserved epidermal growth factor
 RT receptor cDNA from human A431 carcinoma cells.";
 RL Biochem. Biophys. Res. Commun. 124:125-132(1984).
 RN [10]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=88217333; PubMed=3329716;
 RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
 RA Waterfield M.D.;
 RT "The human EGF receptor gene: structure of the 110 kb locus and
 RT identification of sequences regulating its transcription.";
 RL Oncogene Res. 1:375-396(1987).
 RN [11]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=91107677; PubMed=1988448;
 RA Haley J.D., Waterfield M.D.;
 RT "Contributory effects of de novo transcription and premature
 RT transcript termination in the regulation of human epidermal growth
 RT factor receptor proto-oncogene RNA synthesis.";
 RL J. Biol. Chem. 266:1746-1753(1991).
 RN [12]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=85270438; PubMed=2991899;
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
 RT "Characterization and sequence of the promoter region of the human
 RT epidermal growth factor receptor gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
 RN [13]
 RP SEQUENCE OF 540.
 RA Kohda D.;
 RL Submitted (SEP-1997) to the SWISS-PROT data bank.
 RN [14]
 RP RECEPTOR ACTIVITY.
 RX MEDLINE=84191554; PubMed=6325948;
 RA Mroczkowski B., Mosig G., Cohen S.;
 RT "ATP-stimulated interaction between epidermal growth factor receptor
 RT and supercoiled DNA.";
 RL Nature 309:270-273(1984).
 RN [15]
 RP PHOSPHORYLATION.
 RX MEDLINE=89278137; PubMed=2543678;
 RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,
 RA Hawk R., Givol D., Ullrich A., Schlessinger J.;
 RT "All autophosphorylation sites of epidermal growth factor (EGF)
 RT receptor and HER2/neu are located in their carboxyl-terminal tails.
 RT Identification of a novel site in EGF receptor.";
 RL J. Biol. Chem. 264:10667-10671(1989).
 RN [16]

RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
 RP ASN-528.
 RX MEDLINE=96398132; PubMed=8962717;
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;
 RT "Analysis of the glycosylation patterns of the extracellular domain of
 RT the epidermal growth factor receptor expressed in Chinese hamster
 RT ovary fibroblasts.";
 RL Growth Factors 13:121-132 (1996).
 RN [17]
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
 RP ASN-603.
 RX MEDLINE=20198209; PubMed=10731668;
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;
 RT "Characterization of the N-oligosaccharides attached to the atypical
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor
 RT receptor.";
 RL J. Biol. Chem. 273:11150-11157 (1998).
 RN [19]
 RP PARTIAL SEQUENCES, AND DISULFIDE BONDS.
 RX MEDLINE=98225196; PubMed=9556602;
 RA Abe Y., Odaka M., Inagaki F., Lax I., Schlessinger J., Kohda D.;
 RT "Disulfide bond structure of human epidermal growth factor receptor.";
 RL J. Biol. Chem. 273:11150-11157 (1998).
 RN [19]
 RP REVIEW.
 RX MEDLINE=87297456; PubMed=3039909;
 RA Carpenter G.;
 RT "Receptors for epidermal growth factor and other polypeptide
 RT mitogens.";
 RL Annu. Rev. Biochem. 56:881-914 (1987).
 CC -!- FUNCTION: Receptor for EGF, but also for other members of the EGF
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
 CC EGF-like growth factor, Gp30 and vaccinia virus growth factor. Is
 CC involved in the control of cell growth and differentiation.
 CC -!- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
 CC secreted.
 CC -!- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/
 CC truncated isoform/TEGFR, 3/p110 and 4; are produced by
 CC alternative splicing.
 CC -!- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
 CC expressed in ovarian cancers.
 CC -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; X00588; CAA25240.1; -;
 DR EMBL; U95089; AAB53063.1; -;
 DR EMBL; U48722; AAC50802.1; -;
 DR EMBL; U48723; AAC50804.1; -;
 DR EMBL; U48724; AAC50796.1; -;
 DR EMBL; U48725; AAC50797.1; -;
 DR EMBL; U48726; AAC50798.1; -;
 CC -----
 Query Match 44.7%; Score 3028; DB 1; Length 1210;
 Best Local Similarity 48.7%; Pred. No. 5.6e-155;
 Matches 618; Conservative 174; Mismatches 358; Indels 118; Gaps 26;
 QY 11 LLLALLPFGAA--STQVCTGTDKMLRLPASPETHLMDLRHLHYQCQVQGNLELYLPTN 68
 DB 14 LLAALCPASALEKKVCOQTSNKLTLQGTGFEDHFLSLQRMFNNECVVLGNLEITYVQRN 73

QY 69 ASLSFLLQDIQEVQGVYLLIAHNVQVPLQRLRIVRGTLQFEDNYALAVLDNGPLNNTTP 128
 DB 74 YDLSFLKTIQEVAGYVLLIALNTVERIPLENLIQIRGMYVENSVALAVLSNYD----- 126
 QY 129 VTCASPGGLRELOLRSLITELIKGVLIOQNPOLCYODTILWKDIFHNQOLALTLIDTNR 188
 DB 127 ---ANKTGLKELPMRNLQELHGAVFNSNNPALCNVESIQWRDIVSDSDFLSNNMSDFQNH 183
 QY 189 SRACHPCSPMKSGSRGWESSEDCQSILTRTVACAGCA-RCKGPLPTDCHECOCAAGCTGP 247
 DB 184 LGSCKQKDSCPNGSCWGAGEENCQKLTIKIAQCSGRCGRKSPSDCCCHNQCAAGCTGP 243
 QY 248 KHSDDLACILHNHSGICEILHCPALVTYNTDTFSPMPNPEGRTYTFGASCVTACPNYLSTD 307
 DB 244 RESDCLVCRKFRDEATCKDTPPLMLYNTTYQMDVNPESKYSFGATCVKKCPRYVYVD 303
 QY 308 VGSCTLVCPLNQOEVTAEQTCRCEKSKPCARVCYGLNQYIKANSKFGITGILE-FAG 366
 DB 304 HGSCVRACGADSYEM-EEDGVKCKKCEGCRKVCNGIGIGEFK-DLSLSINATNIKFKN 361
 QY 367 CKKIFGSLAFIPESFDGDPASNTAPLQPELOQVETLEETITGYLYTSAMPDLSPLDSVFQ 426
 DB 362 CTSISGDLHLIPVAFRGDSFTHTPDLDPQELDKLVKEITGELLQAWPENTDLHAFE 421
 QY 427 NLQVIRGRILHNGAYSLTLQGLIGISWGLRSLRELGLSLALIHHTHLCFVHTVPMQDLF 486
 DB 422 NLEIIRGRTKHQGFSLAVVSLNITSGLRSLKEISDGDVILSGNKNLCYANTINWKKLF 481
 QY 487 RNPQALLHTANRPEDECYEGELACHQLCARHCWGPGTQCVCNCSQFLRGQECVECRV 546
 DB 482 GTSQKRTKIISNGENSKATGVCHALCSPEGCGPEPRDCVSRNVSREGCDVKCNL 541
 QY 547 LQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHVKDPPFCFVACPCSVKP 606
 DB 542 LEGEPREFVENSICIOCHPECLPQANNITCTGRPDNCIOCAHYIDGPHCVKTCPCAVMG 601
 QY 607 DLSYMPIWKPFDEBEGACQPCINCTHSCVDLDDKGPFAEQRASPLTSIVSAVVG---ILL 663
 DB 602 ENNTL-VWKYADAGHCHLCHPNCTYCTGPGLEGCTNGPKIP--SIATGMVGALLLL 658
 QY 664 VVVLGVVFGILLIKRROOKIRKYMRLLOSTELVEPLTPSGAMPNQOMRILKETELRVK 723
 DB 659 VVALGIG---LFMRRRHIVKRTLRLLQRELVEPLTPSGEAPNOALLILKETEFKI 715
 QY 724 KVLGSGAFGQ----YIKANSKF---IGITELVRENTSPKANKEILDEAYVNAVAGVSPY 775
 DB 716 KVLGSGAFGTVYKGLMPEGEKVKIPVAKI---LREATSPKANKEILDEAYVNAVSDNPH 773
 QY 776 VSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRLGSLQDLNWCMIQAKMSYLEDVRL 835
 DB 774 VCLLIGICLTSTVQLITQLMPFGCLLDYVREHKDNIGSQVLLNWCQVIAKGMVLEDRRL 833
 QY 836 VHRDLAARNVLKSPNHVKITDFGLARLLDIDTEYHADGGKVPKIKMALESILRRPFTH 895
 DB 834 VHRDLAARNVLKTPQHKITDFGLAKLIGAEKEYHAEGGKVPKIKMALESILHRTYTH 893
 QY 896 QSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVIYMLVVKCMI 955
 DB 894 QSDVWSYGVTVWELMTFGSKPYDGIIPASELSSLEKGERLPQPPICITDVIYMLVVKCMI 953
 QY 956 DSECRPRFRELVSFERSMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDGDLVDA 1014
 DB 954 DADSRPKFRELIIIEFSKMARDPQRYLVIQDGRMLHLPSPDTSNFYRALMDEEDMDVDVA 1013
 QY 1015 EYLVFQQGFPCPDPAFGAGVMVHRHSSTRSGGDLTLGLEPSEEEAPRSPAPSEG 1074
 DB 1014 DEYLIPOQGF-----SSPSTRPPLLSSLS 1039
 QY 1075 AGSDVFDGDLGMAAKGLQSLPHTDPSPLORYSEDPVLPSET--DGYVAPLTCPOPE 1132
 DB 1040 ATSN--NSTVACIDRNLGOSCPIKEDSFQORYSSDPTGALTEDSIDDTFL-----PVPE 1091

OY 1133 YVQDPVRPQPPSPREGPLPAARPGATLERAKTLSPKNGVWVDVFAFGAVENPEYL- 1191
 Db 1092 YINQ-SVPKRPAQSVQNPVYHQPLNP-----AFSRDPHYQD--PHSTAVGNPEYLN 1140
 OY 1192 TPQGAAPQHPHPPAFAPFAPFNLYWQO-----DP-----PERGAPPTFFKGTG 1235
 Db 1141 TVQ-----PTCVNSTEDSPAHWAQKGSQHSQSLDNPDYQDFFPKPKAPNGIFKGS- 1190
 OY 1236 TAENPEYL 1243
 Db 1191 TAENAEYL 1198
 RESULT 5
 EGF_MOUSE
 ID_EGFR_MOUSE STANDARD; PRT; 1210 AA.
 AC Q01279;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
 GN EGFR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Liver;
 RX MEDLINE=93026370; PubMed=1408137;
 RA Avivi A., Skorecki K., Yayon A., Givol D.;
 RT "Promoter region of the murine fibroblast growth factor receptor 2
 RT (bek/KGFR) gene."
 RL Oncogene 7:1957-1962(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c, and CD-1; TISSUE=Liver, and Decidua;
 RX MEDLINE=93126380; PubMed=7678348;
 RA Paria B.C., Das S.K., Andrews G.K., Dev S.K.;
 RT "Expression of the epidermal growth factor receptor gene is regulated
 RT in mouse blastocysts during delayed implantation."
 RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Liver;
 RA Hibbs M.L.;
 RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B6/C3; TISSUE=Liver;
 RX MEDLINE=94170986; PubMed=8125255;
 RA Luetsteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
 RA Jenkins N.A., Lee D.C.;
 RT "The mouse waved-2 phenotype results from a point mutation in the EGF
 RT receptor tyrosine kinase."
 RL Genes Dev. 8:399-413(1994).
 RN [5]
 RP SEQUENCE OF 1-714 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=91232866; PubMed=2030916;
 RA Avivi A., Lax I., Ullrich A., Schlesinger J., Givol D., Morse B.;
 RT "Comparison of EGF receptor sequences as a guide to study the ligand
 RT binding site."
 RL Oncogene 6:673-676(1991).
 RN [6]
 RP SEQUENCE OF 969-1117 FROM N.A.
 RC STRAIN=C3H;
 RX Avinger D.P., Serrero G.;
 RA Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
 CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
 CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein

CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.eib-sib.ch/announce/>
 CC or send an email to license@eib-sib.ch).
 CC
 CC EMBL; X78987; CAA55587.1; -
 CC EMBL; U03425; AAA17899.1; -
 CC EMBL; X59698; CAA42219.1; -
 CC EMBL; L06864; AAA53029.1; -
 CC EMBL; Z12608; CAA78249.1; -
 CC HSSP; P11362; IFGK.
 CC MGD; MGI:95294; Egfr.
 CC InterPro; IPR000494; EGFR_L_domain.
 CC InterPro; IPR000719; Euk_pkinase.
 CC InterPro; IPR002174; Furin-like.
 CC InterPro; IPR001245; Tyr_pkinase.
 CC Pfam; PF00069; pkinase; 1.
 CC Pfam; PF00757; Furin-like; 1.
 CC Pfam; PF01030; Recep_L_domain; 2.
 CC ProDom; PD000001; Euk_pkinase; 1.
 CC SMART; SM00261; FU; 3.
 CC SMART; SM00219; TyrKC; 1.
 CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 CC PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 CC Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
 CC Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
 FT SIGNAL 1 24
 FT CHAIN 25 1210
 FT DOMAIN 25 647
 FT TRANSMEM 648 670
 FT DOMAIN 671 1210
 FT REPEAT 75 300
 FT REPEAT 390 600
 FT DOMAIN 1028 1071
 FT NP_BIND 720 728
 FT BINDING 747 747
 FT ACT_SITE 839 839
 FT DISULFID 190 199
 FT DISULFID 194 207
 FT DISULFID 215 223
 FT DISULFID 219 231
 FT DISULFID 232 240
 FT DISULFID 236 248
 FT DISULFID 251 260
 FT DISULFID 264 291
 FT DISULFID 295 307
 FT DISULFID 311 326
 FT DISULFID 329 333
 FT DISULFID 506 515
 FT DISULFID 510 523
 FT DISULFID 526 535
 FT DISULFID 539 555
 FT DISULFID 558 571
 FT DISULFID 562 579
 FT DISULFID 582 591
 FT DISULFID 595 617
 FT DISULFID 620 628
 FT DISULFID 624 636
 FT MOD_RES 680 680
 FT MOD_RES 1092 1092
 CC PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
 CC PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

FT	MOD_RES	1110	1110	PHOSPHORYLATION (AUTO-)	(BY SIMILARITY).
FT	MOD_RES	1172	1172	PHOSPHORYLATION (AUTO-)	(BY SIMILARITY).
FT	MOD_RES	1197	1197	PHOSPHORYLATION (AUTO-)	(BY SIMILARITY).
FT	CARBOHYD	128	128	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	175	175	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	196	196	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	352	352	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	413	413	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	444	444	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	528	528	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	568	568	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	603	603	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	623	623	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CONFLICT	19	19	C -> S (IN REF. 2).	
FT	CONFLICT	539	539	C -> W (IN REF. 5).	
FT	CONFLICT	991	991	L -> F (IN REF. 4).	
FT	CONFLICT	1116	1117	HP -> DR (IN REF. 6).	
FT	SEQUENCE	1210	AA; 134853 MW; 690E20D46DF2D2F5	CRC64;	
Query Match					44.3%; Score 3002; DB 1; Length 1210;
Best Local Similarity					48.6%; Pred. No. 1.4e-153;
Matches 620; Conservative 168; Mismatches 365; Indels 122; Gaps 28;					
QY	11	LLIALLPPGAA--STQVCTGTGDMKRLPASPETHLMDRLHLYQCCQVQGNLEILTYLPTN	68		
Db	14	LLTALCAAGGALEKKVCQGTGSRNLTQGTGTFEDHFLSLQRMNNECEVVLGNLEITYYQVRN	73		
QY	69	ASLSFLQDIOEVQCVYLIHANOVRQVPLORLIRVGTQGLPEDNYALAVLNDGDPNNHTP	128		
Db	74	YDLSFLTKIIEVAGYVLIUANTVERIPLENLQIIRGNALYENTYVALAISN-----	124		
QY	129	VTGASPGGLRELQRLSLEILKGVLIQRPQLCYQDTILMKDT-----FKNNNALTLI	184		
Db	125	-YGNRTGLRELPMRLQELILGAVRESNPNILCNMDTIQWRDIVQNVFMSNMDL---	180		
QY	185	DTNRSRACHPCSPCKSGRSGWESSEDCQSILTRTVACGGCA-RCKGPLPTDCHEQCAAG	243		
Db	181	-QSHPSPCRCDESPCNGSGWGGEECQKLTIKCAQCQSHRCRGRSPSDCCNQCAAG	239		
QY	244	CTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCYTACPVNY	303		
Db	240	CTGPRESDCLVCKQFQDEATCKDTCPPLMLYNTTYQMDVNPPEGYSFGATCVKCKCRNY	299		
QY	304	LSTDVSGCTLVCPLHNOEVTAEQDTCRCCKSPCARVCYGLGQYIKANSKFTIGITELE	363		
Db	300	VTDHSGSVACAGPDYEV-EEDGIRKCKDGCRCVKNGIGIGEFK-DTLSINATNIK	357		
QY	364	-FAGCKIIFGLAPLPESFGDGPASNTAPLOPELOVFTELEHTGYLYISAWPDSLPDL	422		
Db	358	HFKYCTAISGDHLPLVAFKGSFTRTPPLDPRELEILTKVKEITGFLLIQAWPDNWTDL	417		
QY	423	SVFONQVIRGRIHNGAYSITLQGLGSIWGLRLSRELGLALIHNTHLCEVHTVPW	482		
Db	418	HAFENLEIRGRTKHQGFQSLVAVGLNITSLGRSLKEISDGDVITISGNRLCYANTINW	477		
QY	483	DQLFRNPQALLHTANPEDECVGEGIACHCARGCWGPGTQCVCNCSOFLRGOCVE	542		
Db	478	KKLFGTGNQTKIMNNAEKDKAVHNVCPNCSSECGCWGPEPRDCVQCNVSGRECV	537		
QY	543	ECRVQLGPREYNARHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFPCVACRPS	602		
Db	538	KCNILEGEPREFENSECIIQCHPECLPQAMNITCTGRGPDNCTQCAHYIDGPHCVKTCPA	597		
QY	603	GVPKDLSPYMPKPPDEEGACQCPINCTHSCVDLDDKGCPEAQRASPLTSIYSAVIGIL	662		
Db	598	GIMGENNLT-LVKYADANNVCHLCHANCTYGCAGPGLQGCVEWPSGPKIESIATIGVGL	656		
QY	663	LVVVLGVVFGI-LIKRQKIRKTYMRLLQVELVEPLTPSGAMPNQAMRILKETEELR	721		
Db	657	LFIVV-VALGIGLFMRRHVRKTLRLQLQRELVEPLTPSGEAFNQAHRLIKETEFLK	715		
QY	722	KVKVLGSGAFQ-----YIKANSKF-----IGITELVLRENTSPKANKEILDEAYVMAGVS	773		

Db	716	KIVLGSAGFETVYKGLWIPGEKVKIPVAIKE--LREATSPKANKEILDEAYVMASVDN	773		
QY	774	PYVSRLLIGICLTSTVOLVTOLMPYGCLLDHHVNRGRGLSGQDGLNWCMIKAGMSYLEDV	833		
Db	774	PHVCRLLIGICLTSTVOLITOLMPYGCLLDYVRHKNIGSQYLLNWCMIKAGMSYLEDR	833		
QY	834	RLVHRDLAARNVLKSPNVHKITDFGLARLIDDETEYHADGKGKVPKMWMALESILRRRF	893		
Db	834	RLVHRDLAARNVLKTPQHVKITDFGLAKLLGAEKEEYHAEKGVKPIKMWMALESILHRIY	893		
QY	894	THOSDVMSYGVTVWELMTFCAGKPYDGPAREIPDLEKGERLIPQPPCTIDVYMWKWC	953		
Db	894	THOSDVMSYGVTVWELMTFGSKPYDGPASDISILEKGERLIPQPPCTIDVYMWKWC	953		
QY	954	MIDSECRPRFRELVSFMRMARDPQRFVVIQ-NEDLGASPLDSTFYRSLEDDMDGDLV	1012		
Db	954	MIDADSRPRFRELILFESKWARDPQRYLVIOGDERMHLFPTDSNFYRALMDEDMEDVV	1013		
QY	1013	DAEYILVPOQGFPCPDPAFGAGGMVHHRSSSTRSGGGDLTLGLEPSEEEAPRSLAPS	1072		
Db	1014	DADEXLIPQCGFF-----NSPST-----SRTPLSS	1039		
QY	1073	EGAGSDVFDGLMGAAKGLQSLPTHDPSPLOQRYSEDPTVPLPSET--DGYVAPLTCSPQ	1130		
Db	1040	LSATSN-----NSTVACINRNGSCRVEDAPLQRYSSDPTGAVTEDNIDDAFL-----PV	1089		
QY	1131	PEVYNOPDVRFPQPPSPREGPLPAARAGATLERAKTILSPGKNGVVKDVFAGGAVENPEY	1190		
Db	1090	PEVYNO-SVPRPAGSVQNPVYHNPDPH-----APGRDLHYQN--PHSNAVGNPEY	1138		
QY	1191	L-TPQGAAPQHPHPPAFADNLVYWDQ-----DP-----PARGAPPSTFKG	1233		
Db	1139	LNTAQ-----PTCLSGGFSNPAWIKQSHQMSLQNDPQDFFPKETKENGIFKG	1189		
QY	1234	TPTAENPEYGLDVP 1248			
Db	1190	-PTAENAEYLRVAPP 1203			
RESULT 6					
ERB4_HUMAN	ERB4_HUMAN	STANDARD;	PRT; 1308 AA.		
AC	Q15303;				
DT	15-DEC-1998 (Rel. 37, Created)				
DT	15-DEC-1998 (Rel. 37, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)				
DE	(p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).				
GN	ERB4 OR HER4				
OS	Homo sapiens (Human)				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A. (ISOFORM JM-A).				
RC	TISSUE=Breast carcinoma;				
RX	MEDLINE=93189574; PubMed=8383326;				
RA	Plowman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,				
RA	Foy L., Neubauer M.G., Shoyab M.;				
RT	"Ligand-specific activation of HER4/p180erbB4, a fourth member of the				
RT	epidermal growth factor receptor family.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).				
RN	[2]				
RP	SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).				
RC	TISSUE=Fetal brain;				
RX	MEDLINE=97476287; PubMed=9334263;				
RA	Elenius K., Corfas G., Choi C.J., Rio C., Plowman G.D.,				
RA	Klagabrun M.;				
RT	"A novel juxtamembrane domain isoform of HER4/erbB4. Isoform-specific				
RT	tissue distribution and differential processing in response to				
RT	phorbol ester.";				
RL	J. Biol. Chem. 272:26761-26768(1997).				


```

FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 496 633 CYS-RICH.
FT DOMAIN 718 985 PROTEIN KINASE
FT NP_BIND 724 732 ATP (BY SIMILARITY).
FT BINDING 751 751 ATP (BY SIMILARITY).
FT ACT_SITE 843 843 BY SIMILARITY.
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 246 BY SIMILARITY.
FT DISULFID 249 258 BY SIMILARITY.
FT DISULFID 262 289 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.
FT DISULFID 308 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 559 577 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 614 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 621 633 BY SIMILARITY.
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 1062 1062 S -> N (IN REF. 3).
FT CONFLICT 1080 1082 PYT -> SYR (IN REF. 3).
SQ SEQUENCE 1308 AA; 146957 MW; D944BB0996A08B41 CRC64;

Query Match 41.8%; Score 2830; DB 1; Length 1308;
Best Local Similarity 43.9%; Pred. No. 2.5e-144;
Matches 591; Conservative 191; Mismatches 399; Indels 166; Gaps 29;

Qy 1 MELA-ALCRWGLLL--ALLPPGAASCTVCTDMKRLPASPETHLDMLRLHYOGCQVQ 57
Db 1 MKLATGLVWGSLLVAARTVPSASQSVACATENKLSLSLEQQYRALRYNCEVVM 60

Qy 58 GNLELTLYPTNASLFLQIDIOGVYVLIHNOVRQVPLQRLRIVRGTLQFEDNYALAVL 117
Db GNLEITSIEHNRDLNFLRSIREVTGYVLVALNQRYLPLENLRIIRGTLKLYEDRYALAIF 120

Qy 118 DNGDPLNNTTPTVGASPGGLRELOIRSLTEILKGVLIQRPQLCYQDTILWKDIFHNN 177
Db 121 LNYRKDGNF-----GLOELGNLTILNGVYVYDQNKFLCYADTIHWQDIVRNPW 171

Qy 178 QLALTLIDTNRSRACHPCSPCKSRGWSSESDECOQLTRTVACAGGC-ARCKGPLPTDCC 236
Db 172 PSNMTLVSTIGSSGGRCHKCTG-RCWGPTEHCOQLTRTVACAEQDCGRYGVVSDCC 230

Qy 237 HEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFSPMPNPEGRYTFGASCV 296
Db 231 HRECAAGCGSPKDTDCFACMNFNDSGACVTCQPTFVYVNTPTTFQLEHNFNAKYTYGAFV 290

Qy 297 TACPYNLYSTDVGSCTLVCPHNOEVTAEADTQRCCKSKPCARVCVGLNMQYIKANSKF 356
Db 291 KCCPHNFV-VDSSSCVRCAPSSKMEV-EENGIRKMKCPCTDICPKACDGTGIGTSLMSAQTV 348

```

RESULT 8
XMRK_XIPMA
ID XMRK_XIPMA
AC P13388;

STANDARD; PRT; 1167 AA.

```

Qy 357 IGITELEFFAGCKKIFGSLAFPLPESFDDPASNTAPLQEQLOVETLEETITGLYISAMP 416
Db 349 DSSNIDKFNCTKINGNLIFLVTGIHGDYPNAIDDAIDPEKLVNFRVITGTFLNIQTWP 408

Qy 417 DSLPDLVSFONLQVIRGRILHNGAYSILTLOGLGISWLGSLRSLBELSGSLAIHHTLCLF 476
Db 409 PNMTFDSFVSNLVTIGRVLVSLGLSLILKQQGITSLOFQSLKEISAGNTIYITDNSLCY 468

Qy 477 VHTVPMDQLFRNPHQALLHTANPEBECVGECHQACARGHCWGPPTQCVNCOSFLR 536
Db 469 YHTINWTLFSTVNRQIRIVDRNRRAENCTAEGVWCHLCSNDCGWCWGPDPQCLSCRFSR 528

Qy 537 GQECVEECRYLOGLPREYVNAHCLPCHPECQ- QNGSVTCFGEADQCVCACAHYKDPFF 595
Db 529 GKICIESCNLYDGEFREFENGSIQVCDSCQERKEDGLTCHGPGDNCCTKCSHFKDGPN 588

Qy 596 CVARCPGVKPDLSYMPIWKFPDEEGACQCPINCTHSCVDLDDKGC-----PA 644
Db 589 CVEKCPDVLQGANSF--IFKYADQDRCHPCHPNCTQGCNGPTSHDCIYYPWTGHSFLPQ 646

Qy 645 EQRASPLTSTVSADV--GILLVVLGVVFGILIKRROQKIRKYTMRLLOETELVEPLTPS 703
Db 647 HAR-TPL--TAAGVIGGLFLVIMALTFAVYVRKSIK-KKRALRPL-ETELVEPLTPS 701

Qy 704 GAMPNQAQMRILKETELRKVKVLGSGAFQYIK-----ANSKFIGITELVLRENTSPKA 757
Db 702 GTAPNQAQRLILKETELRKVKVLGSGAFGVYKGIWVPEGETVKIPIVAIKILNETTGPKA 761

Qy 758 NKEILDEAYVMAVGSPYSRLLGICLTSTVQLVLTQMPYGLLDHVRNRLGSGDILL 817
Db 762 NVEFMDREALTMASVDHPHLVRLGLVCLSPITQLVTLMPHGCILLEYVHEHKDNGISOLL 821

Qy 818 NWCWQIAGKMSYLEVRLVHRDLAARNVLKSNHVKITDPLGLARLIDIDETEVHADGGK 877
Db 822 NWCQIAGKMSYLEVRLVHRDLAARNVLKSNHVKITDPLGLARLIDIDETEVHADGGK 881

Qy 878 VPIKMWALLESILARRRTHOSDVMYSYGVVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQ 937
Db 882 MPIKMWALECIHYRKETHQSDVMSYGVVWELMTFGKPYDGIPTREIPDLLEKGERLPQ 941

Qy 938 PPICTIDVYMWKMWIDSECPRELVSFESRMARDPQRVFIQNEQ-LGPASPLDS 996
Db 942 PPICTIDVYMWKMWIDADSRPKELAAEFSSRMARDPQRYLVITQGGDRMKLPSNDS 1001

Qy 997 TFYRSLEDDDDGDLVDAEYLVPOQGFCDP-----APGA 1033
Db 1002 KFFQNLLEDDLEDDMDAEYLV-PQAFNIPPIYTSRTRIDNSRSEIGHSPPPAYTPMS 1060

Qy 1034 GGMVHHRHRSSTRSGGDLTLGLEPSEEBEAPRSLAPSEAGSDVFDGLGMAAKGLQ 1093
Db 1061 GSQFVYQDGGFATQGG---MPMPYTATTSTIPEAPVA--QGATAEMFDDSCCNGTLRKPV 1115

Qy 1094 SLTHDPSPLOQRYSEPTVPLPS-----ETQGVVAPLTCSPQPEYVNOVDVRRPQPPSP 1146
Db 1116 VPHVQSDSSQRYASDPTVPAPERNPRAELDEEGYMTPMHDKPKQBYLNPVE-----1167

Qy 1147 REGPLPAARPAAGATLERAKTLSPGKGVKVDVAFAGGAVENPEYLTPOGGAAPQHPPPPA 1206
Db 1168 -ENPFVSSRR-----KNGDLQ-----ALDNPETHSASSG-----PPKA 1198

Qy 1207 -----FSPAFDNLVYWDQDPPPERGA--PPST 1230
Db 1199 EDEYVNEPLYLNTFTNALGNAEYMKNSLLSVPEKAKKAFDNPYWNHSLPRSTLQHPDY 1258

Qy 1231 FKGTPT-----AENPEYL 1243
Db 1259 LOEYSTKYFYKONGRIRPIVAENPEYL 1285

```

DT 01-JAN-1990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
GN XMRK OR TU.
OS Xiphophorus maculatus (Southern platyfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthopterygii; Acanthopterygii; Percormorpha; Atherinomorpha;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX NCBI_TaxID=8083;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90015140; PubMed=2797166;
RA Wittbrodt J., Adam D., Malitschek B., Maueuer W., Raulf F.,
RA Telling A., Robertson S.M., Schartl M.;
RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
RT inducing Tu locus in Xiphophorus";
RL Nature 341:415-421(1989).
RN [2]
RP REVISION TO 515.
RA Schartl M.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; X16891; CAA34770.2; -.
DR PIR; S06142; IFGK.
DR HSP; P11362; IFGK.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR ProDom; PD01030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.
KW Transmembrane; Glycoprotein; Receptor; Signal; Transferrase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
FT SIGNAL 1 25
FT CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE
FT KINASE.
FT DOMAIN 26 642
FT TRANSMEM 643 665 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 666 1167 POTENTIAL.
FT DOMAIN 710 977 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 716 724 PROTEIN KINASE.
FT BINDING 743 743 ATP (BY SIMILARITY).
FT ACT_SITE 835 835 ATP (BY SIMILARITY).
FT DISULFID 195 204 BY SIMILARITY.
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 220 228 BY SIMILARITY.
FT DISULFID 224 236 BY SIMILARITY.

FT DISULFID 237 245 BY SIMILARITY.
FT DISULFID 241 253 BY SIMILARITY.
FT DISULFID 256 265 BY SIMILARITY.
FT DISULFID 269 296 BY SIMILARITY.
FT DISULFID 300 311 BY SIMILARITY.
FT DISULFID 315 330 BY SIMILARITY.
FT DISULFID 333 337 BY SIMILARITY.
FT DISULFID 504 513 BY SIMILARITY.
FT DISULFID 508 521 BY SIMILARITY.
FT DISULFID 524 533 BY SIMILARITY.
FT DISULFID 537 553 BY SIMILARITY.
FT DISULFID 556 569 BY SIMILARITY.
FT DISULFID 560 577 BY SIMILARITY.
FT DISULFID 593 615 BY SIMILARITY.
FT DISULFID 618 626 BY SIMILARITY.
FT DISULFID 622 634 BY SIMILARITY.
FT CARBOHYD 114 114 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 144 144 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 201 201 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 356 356 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 365 365 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 398 398 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 621 621 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;
Query Match 37.9%; Score 2569.5; DB 1; Length 1167;
Best Local Similarity 44.0%; Pred. No. 2e-130;
Matches 558; Conservative 166; Mismatches 399; Indels 145; Gaps 30;
QY 4 AALRCGLLIALPPGAAS ---QVCTGDMKRLPASPETHLDMRLHLYOGQVVOGN 59
DB 8 AALLQ--LLLVLSISCCSTDPDRKVCQGTSNQTM---LDNHYLKKMKYSGCNVLEN 62
QY 60 LEITYLPTNASLFLQDIQEVQGYVLIANHNVQVPLQRLRIRVGTQLFEDNYALAVLDN 119
DB 63 LEITYTQENQLSFLQSIQEVGGYVLIAMNEVSTIPLVNLRLIRGQNLVEGNTFLVMSN 122
QY 120 GDPNNNTTPVTGASPGRLRELQRLSITELKGVLIQRPOLCYQDTILWKDIFHKNQL 179
DB 123 YQK-NPSSP--DYVQVGLKQLQLSNITELSGGVKVSHPNLLCNVETINWMDIVDKTSNP 179
QY 180 ALTLIDTNRSRACHPCSPMKSGRSGESSEDQSLTRTVCAGC-ARCKGPLPTDCCE 238
DB 180 TNNLIPIAFERQCKQDHGCVGSCWAPGPGHCQKFTKLLCAEQCNRCRGPIDCCNE 239
QY 239 QCAAGCTGPKHSDCLACLFPHNHSGLCELHCPALVTYNTDTFESMPNPEGYTTFGASCVTA 298
DB 240 HCAGGCTGPRATDCLACRDFNDGTCCTCPPPKIYDIVSHQVVDVNPNIKYTFGAACVKE 299
QY 299 CPYNYLSTDVGSCTLVCPPLHNQVETAEDGTORCEKSKPCARVCYGLGMOYIANKSFIG 358
DB 300 CPSNYVYTE--GACVRS CSAGMLEVD--ENGRSKCPDGVCPKVDGIGISL--SNTIAVN 356
QY 359 ITBL-EFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQVFTLEETIYGLVISAQPD 417
DB 357 STNIRFSNCTKINGDIILNRNSFEGDPHYKGTMDPEHLWNLTVTKEITGYLVIMWNP 416
QY 418 SLPDLVFNQLQVIRGRIHLNGAYS-LTLQGLGISWLGLRLSLRSLRELGLSLIHHNTHLCF 476
DB 417 NMTLSLVFQNLLEIIRGTRTTFSGRGSFVVQVVRHLQMLGLSLKEVSAGNVILKNTLQRY 476
QY 477 VHTVPMDQLFRNPQALLHTANRPEDECVGEGLACHOLCARGHCWGPGTQCVCNCSQFLR 536
DB 477 ANTNIRRLFRSEDSQSI EYDART-----ENQTCNNECEDGCGWPGTMCVSCSLHVD 529
QY 537 GQSCVEECVQLGLPREYVNAHCLPCHPECPQNGSVTCFCGPEADQCVACAHYKDPFPC 596
DB 530 GGRCVASCNLLQGEPREAQVDGRCVQCHOECLVQTSLSLCYGPGPANCKSAHFQDGPQC 589
QY 597 VARCPSGVKPDLISYMPIWKFPFDEEGACQPCPINCTHSCVDLDDKGCPCBQASPLTSIVS 656

Db 590 IPRCPHGILGSDTL-IWKYADKMGQCPCHONCTQCGSGPLSGCRGD-IVSHSLAVG 647
Qy 657 AVGILLVVLGVVFGVGLIKKQIRKYTWRRLOETELVEPLTPSGAMNQOMRLK 716
Db 648 LVSGLLITVIVALLIVLRRRIK-RKRTIRCLQEKELVEPLTPSGAQNQAFRLK 706
Qy 717 ETELKVKVLGSGAFGOYIK-----ANSKFIGITELVRLRENTSPKANKEILDAYVAG 770
Db 707 ETEFKORVLGSGAFGVYKGLWNPDCENIRIPVAKVLRATSPKYNQEVLDAYVAS 766
Qy 771 VGSYVSRLLGICLTSTVQLVQLMPYGCCLLDHVRNRLGSLDNLNWCQIAKMSYL 830
Db 767 VDHPRVCRLLGICLTSAVLVQLVQLMPYGCCLLDYVROHQERICGQWLLNWCQIAKMSYL 826
Qy 831 EDVPLVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIKWMALESILR 890
Db 827 EERHLVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIKWMALESILQ 886
Qy 891 RRFTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEGERLPPOPICTIDVYIMV 950
Db 887 WTYTHQSDVMSYGVTVWELMTFGSKPYDGPAREIPDLLEGERLPPOPICTIEVYIL 946
Qy 951 KCMWIDSECRPRFRELVSFMRARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGD 1010
Db 947 KCMWIDPSSRPRFRELVSFMRARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGD 1001
Qy 1011 LVDAEVLVPOGFFCPDPAPGAGMVHRRHSSTRSGGDLTLGLLEPSEEEAPRPLA 1070
Db 1002 VVDAEVLVPOGFFCPDPAPGAGMVHRRHSSTRSGGDLTLGLLEPSEEEAPRPLA 1024
Qy 1071 PSEGAGSDVDFDGLGMAAGKLSLTHDPSPLQRYSEDPTV-PLPSETDGYVAPLTCSP 1129
Db 1025 PPTGH-----PVRENSITLRLNSDPTQNALEKDLGH----- 1056
Qy 1130 QPEYVNPQVVRQP-----PSPRE-----GLP-AARPAATLERAKTLSPCKNGVVK 1176
Db 1057 --EYVNPQGETSSRLSDIYNPNVEDLTDGWPVSLSSQEAETNFSRPEVLTNQNSL-- 1112
Qy 1177 DVFAFGAVENPEYLTQGGAPQHPPPAPSPAFDNLVYDQDPPERGAPPSTFKCTPT 1236
Db 1113 -PLVSSGMDPDY---QAG-----YQAAF-----LPQTGALTNGMFLPA 1149
Qy 1237 AENPEYLG 1244
Db 1150 AENPEYLG 1157
PRT; 1342 AA.
STANDARD;
ID_ERB3 HUMAN
AC P21850;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERB3 OR HER3
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2687875;
RA Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERB3, a third member of the
RT ERB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors."
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;

RA Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene."
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN (3)
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Placenta;
RX MEDLINE=93282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase."
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M29366; AAA35790.1; --
CC EMBL; M34309; AAA35979.1; --
CC EMBL; S61953; AAB26935.1; --
CC PIR; A36223; A36223.
CC HSP; P11362; IFGK.
CC Genew; HGNC:3431; ERBB3.
CC MIM; 190151; --
CC InterPro; IPR000494; EGFR_L_domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF01030; Recep_L_domain; 2.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 3.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
CC PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
CC Alternative splicing.
FT SIGNAL 1 19
FT CHAIN 20 1342
FT DOMAIN 20 643
FT TRANSMEM 644 664
FT DOMAIN 665 1342
FT DOMAIN 709 966
FT NP_BIND 715 723
FT BINDING 742 742
FT ACT_SITE 834 834
FT DISULFID 186 194
FT DISULFID 190 202
POTENTIAL.
RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
EXTRACELLULAR (POTENTIAL).
POTENTIAL.
CYTOPLASMIC (POTENTIAL).
PROTEIN KINASE.
ATP (BY SIMILARITY).
ATP (BY SIMILARITY).
BY SIMILARITY.
BY SIMILARITY.

RC STRAIN=Sprague-Dawley; TISSUE=Liver;
 RX MEDLINE=96096535; PubMed=8522190;
 RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
 RT "Cloning of the rat ErbB3 cDNA and characterization of the
 RL recombinant protein.";
 RL Gene 165:279-284 (1995).
 [2]
 RP REVISIONS TO 85; 513 AND 565.
 RA Hellyer N.J., Koland J.G.;
 RL Submitted (DSC-2001) to the EMBL/GenBank/DBJ databases.
 [3]
 RN SEQUENCE OF 922-1097 FROM N.A.
 RP STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
 RC MEDLINE=97184212; PubMed=9030624;
 RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
 RT "Expression of neurotrophins and their putative receptors, ErbB2 and
 RL ErbB3, is induced during Wallerian degeneration.";
 RL J. Neurosci. 17:1642-1659 (1997).
 CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
 CC -1- LIGAND: BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
 CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
 CC PHOSPHATIDYLINOSITOL 3-KINASE.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U29139; AAC28498.2; -;
 DR EMBL; U52530; AAC53050.1; -;
 DR HSSP; P11362; IFGK.
 DR InterPro; IPR000494; EGFR_L domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 5.
 DR SMART; SM00219; TyrcK; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; FALSE NEG.
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
 DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Tyrosinase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 19
 FT CHAIN 20 1339
 FT DOMAIN 20 643
 FT TRANSMEM 644 662
 FT DOMAIN 663 1339
 FT DOMAIN 183 259
 FT DOMAIN 707 964
 FT NP_BIND 713 721
 FT BINDING 740 740
 FT ACT_SITE 832 832
 FT DISULFID 186 194
 FT DISULFID 190 202
 FT DISULFID 210 218
 FT DISULFID 214 226
 FT DISULFID 227 235

FT	DISULFID	231	243	BY SIMILARITY.
FT	DISULFID	246	255	BY SIMILARITY.
FT	DISULFID	259	286	BY SIMILARITY.
FT	DISULFID	290	301	BY SIMILARITY.
FT	DISULFID	305	320	BY SIMILARITY.
FT	DISULFID	323	327	BY SIMILARITY.
FT	DISULFID	500	509	BY SIMILARITY.
FT	DISULFID	504	517	BY SIMILARITY.
FT	DISULFID	520	529	BY SIMILARITY.
FT	DISULFID	533	549	BY SIMILARITY.
FT	DISULFID	556	573	BY SIMILARITY.
FT	DISULFID	576	585	BY SIMILARITY.
FT	DISULFID	589	610	BY SIMILARITY.
FT	DISULFID	613	621	BY SIMILARITY.
FT	DISULFID	617	629	BY SIMILARITY.
FT	CARBOHYD	126	126	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	250	250	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	353	353	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	408	408	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	414	414	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	437	437	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	469	469	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	522	522	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	566	566	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	616	616	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	1028	1028	L -> P (IN REF. 3).
SQ	SEQUENCE	1339	AA; 147545	MM; 0AA5F2402BBDFIE CRC64;

Query Match 33.1%; Score 2241.5; DB 1; Length 1339;
 Best Local Similarity 39.8%; Pred. No. 8.6e-113;
 Matches 512; Conservative 168; Mismatches 440; Indels 165; Gaps 36;

QY	3	LAALRWGLLLALLPPGAA---STVCTCTDMKLRLPASPETHDMLRLHLYQCQVVQGN	59
DB	7	LQVLC-----FLSLARGSEWNGNSQAVCPGLTNGLSVTGDADNOYQTYLYKLYEKEVVMGN	62
QY	60	LEITLPTNASTLSFLQDIQEVQGYVLIHNOVRQVPLQRLRIVRGTQQLFEDNTALAVLDN	119
DB	63	LEIVLTGHADLSFLQWIREVTGYVLVAMNEFSLPLPLNLVRVGTQVYDGKFAIFW--	120
QY	120	GDPLNNTPTVGTASPGGLRELQRLSRLTEILKGVLTORNPOLCYQDTILMKDIFHKNNQL	179
DB	121	---LNTNT---NSSHALROLKFTQLTEILSGGVYIEKNDKLMCHMDTIDWRDVRVR--	170
QY	180	ALTLDITNRSRACHPCSPMKSGRCSSESDECSLTTRTVACGGC-ARCKGPLPTCCHE	238
DB	171	GAEIVVKNNGANCPPCHEVCCKG-RCWNGPGDDCQILTKTKICAPQCNGRCFNPNCQCHD	229
QY	239	QCAAGCTGPKGSHDCLALHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA	298
DB	230	ECAGCGSGQDITDCFACRRFNDSGACVPRCPPELVYVKNLTFLQLEPNHTKYQYGGVCVAS	289
QY	299	CPYNVLTDSVGSCTLVCPHLNQBVTABDGTQRCCKSKCPARVCYGL--GMOYIKANSKF	356
DB	290	CPHNFV-VDQTFVCRACPPDKMEVD-KHGLKMCPCGGLCPKACEGSGSRQYTVDDSN	347
QY	357	IGITELEFACCKIFGSLAFLPESFGDPSASNTAPLQPEQLQVFEETLEETGVLISAWP	416
DB	348	ID----GFVNCTKILGNLDFLITGLNVDPMHKIPALDPEKLVNFRVTRIVTGLYNTQSWP	403
QY	417	DSLPLDSVFONLQVIRGRILHNGAYS-LTIQIGISWLGSLRLSRLSGSLALHNNTHLC	475
DB	404	PHMNFVSFSLNLTIGRSLYNGRFSLLMKNLNVTSLGFRSLKETLSAGRVYISANQQLC	463
QY	476	FVHTVPWDLFRNPHOALLHTA-NRPEDECVGEGLAHQLCARHGCHWGPGPTQVNCNQSF	534
DB	464	YHSLNWTLLRGLSPSERLDIKYDRPLGLECLABGKVCPLCSGGCGWPGPGQCLSCRNY	523
QY	535	LRQECVEECRVLOGLPREVYNARHCLPCHPECPQNGSVTCFEPADQCVACAHYKDP	594
DB	524	SREGVCVTHCNFLQGBEPREFVHEAQCFSCHEPCEPMEGTSTCNGSGSDACARCAHFRDGP	583
QY	595	FCVARCPGKVPDLSYMPIWKFPDEEGACQPCPCINCTHSC--VDLDDKGPAPQASPLT	652

Db 584 HCVNSCPHGLG--AKGIYKYPDAQNECRPCHENCITQCGNPELQDCLGQAEVLMSPKH 641
Qy 653 SIVSAVVGILVAVGVVFGILIKRRQOKIR-KYTMERLLQETELVELPTPSGAMPNOAQ 711
Db 642 LVIAVTVG--LAVILMILGGSFLYWRGRIONKRAMRYLGRGESIFLPDS-EXANKVL 698
Qy 712 MRILKETELRKVKVLGSGAFQYIK-----ANSKFTIGITELVURENTSPKANKEILDEA 765
Db 699 ARIFKETELRKVLGSGVFGVTHKGIWIPGESIKIPVCIKVIEDSGRQSQFQAVTDMH 758
Qy 766 YVMAGVSPVYRSLGLCTSTVOLTPMYPGCLLDHVRENRCRLGSDLLWCMQOIAK 825
Db 759 LAVGSLDHAHIVRLGLCPGSSLGVLYPLGSLLDHVHVKHRTGLPQLLNWGVQIAK 818
Qy 826 GMSYDELVRHLDAARNVLKSPNVHKITDFGLARLLDIDEYHADGGKVPKMMAL 885
Db 819 GMYLEHSHVHRDLARNVLMKSPSQVQVADFGVADLLPDDKQLLHSEAKTPIKMMAL 878
Qy 886 ESILRRRFTQSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDV 945
Db 879 ESIHFGKYTHQSDVMSYGVTVWELMTGABPYAGLRLAEIPDLLEKGERLAQPOICTIDV 938
Qy 946 YMIWVKCWMIDSECRPRFELVSEFSRWARDPQRFVVIQNELDGPASPLDSTFVRSILED 1005
Db 939 YMWVKCWMIDENIRPTFKELANEFTRWARDPPRYLVIKRAS-GPGTP--PAAEPSVLTT 995
Qy 1006 DMDGLVDABEYLVPQGFPCDPAPGAGGVVHRRHSSSTRSGGDLTLGLEPSES-- 1062
Db 996 KEL-----QEALEPEL-----DLDLLEAESEGLA 1021
Qy 1063 -----EAPRSIAPSEG-----AGSDVFDGLGMGAAGLQSLPT 1097
Db 1022 TSLGSALSLEPTLTLPGRGSSLLSPSSGYMNMNQSSILGEACLDLSAVLGGREQPSRISL 1081
Qy 1098 HDPSPLQRYSEDPTVPLPSETDGVV----APL-----TC-----SPOPE-----YVNPQ 1137
Db 1082 H-PIPRGR-----PASESEGHVTSGEAELOKSVCRSRSRSPRGRGDSAYHSQR 1133
Qy 1138 DVRPQPPSPREG-----LPAARPAGATLERAKLTSP-GKGVV-----KDVVF 1179
Db 1134 HSLTPTVPLSPGLEBEDNGYVMPDTHLRGASSREGTILSSVGLSVLGTDEED-- 1191
Qy 1180 AFGGAVENPEYLTPOGGAAPQHPHP 1204
Db 1192 -----EEYVYMKRRGSP-PRPP 1209

RESULT 11

EGFR DROME STANDARD; PRT; 1426 AA.
AC P0412; Q61601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DE (Gurken receptor) (Torpedo protein) (Drosophila relative of ERBB).
GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS Drosophila melanogaster (Fruit fly)
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RP SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX MEDLINE=94350209; PubMed=8070664;
RA Clifford R., Schubach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT that several genetically defined classes of alleles cluster in
RT subdomains of the receptor protein."
RL Genetics 137:531-550(1994).
RN [2]
RP REVISTONS.

RA Clifford R., Schubach T.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Livneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RT hormone binding and kinase domains.";
RL Cell 40:599-607(1985).
RN [4]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC STRAIN=Oregon-R, TISSUE=Embryo;
RX MEDLINE=87002474; PubMed=3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RT Drosophila EGF receptor homolog transcripts.";
RL Cell 46:1091-1101(1986).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
RT specification in wild-type, Ellipse, and null mutant Drosophila.";
RL Dev. Biol. 205:129-144(1999).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Avril J.F., Agayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos F.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flesler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapieton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [7]
RP SEQUENCE OF 959-1078 FROM N.A.
RC STRAIN=Daekwanyong;
RX MEDLINE=85137938; PubMed=2983232;
RA Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
RT "A Drosophila genomic sequence with homology to human epidermal


```

Db 603 IADCGYISNAYK--FDNRKTKICKPECK-----TCNGAGADHCOEQRVHRDQGHQVSEC 654
Qy 601 P-----SGVK-----PDL 608
Db 655 PKNYNDRGVCRECHATCGTGKPTIGIGACTTCNLAIIINNDATVKRCLLKDKCPD- 713
Qy 609 SYMFIWKF--PDEGACQP-----CPI-----NCH- 632
Db 714 GY--FWYVHPQEOGSLKPLAGRAVCRKCHPLCLCTNYGYHEQVCSKCTHYKREQCET 771
Qy 633 -----SC-----VLDLDDG----- 641
Db 772 ECPADHTDEBQRCFORHPHCNCTGPGADDCKSRNFKLFANETGPGYVNMFMNCTS 831
Qy 642 -CPAEQR-----ASPLTS-----IVSAVVGILLVVLGVVFGI 673
Db 832 KCPLMRHVNQYTAIGPYCAASPPRSKITANLDVNMFIITGAVLVPTICILCVV--T 889
Qy 674 LKRRQOKIRKYT--MRRLQOETELVEPLTPSGAMPNQAOQMRILKETELRKVKVLGSGAF 731
Db 890 YICRQOKAKETVKMTMALSGCEDSEPLRPSNIGANLCKLRIVKDAELRKGGLGNGAF 949
Qy 732 GOYIKA-----NSKF-IGITELVLRNTSPKANKELDEAYVMAGVGSPPYSRLIGIC 783
Db 950 GRVYKGVWPEGENVKIPVAIKELL--KSTGAESSEBFLREAYIMASEEHNLLKLAVC 1007
Qy 784 LTSTVQLVTOLMPYGCILLDVRNRRGLSGODLLNWCQIAKMGSYLEDVRLVHRDLAAR 843
Db 1008 MSSQMLLITQLMPLGCLLDYVRNRRDKIGSKALLNWSQTAKGMSYLEEKLVRDLAAR 1067
Qy 844 NVLVKSNHVKITDFGLARLLIDDETYHADGGKVPKIMMALESILRRRTHQSDVMSYG 903
Db 1068 NVLVQTPSLVKITDFGLAKLLSSDSNEYKAAGGKWPVKWLALEICIRNRVFTSKSDVWAF 1127
Qy 904 VTVWELMTFCAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYMWKCMWIDSECEPRF 963
Db 1128 VTIWELLTUGORPHENIPAKDIPLIEVGLKLEQPEICSLDIYCTLLSCHWLDAMREPTF 1187
Qy 964 RELVSEFSRMARDPQRFWIQLNEDLG--PASPLDSTFYRSLLEDD--DMGDLVDABEYL 1018
Db 1188 KQLTTFVAFEFARDGRIYALPGDKFTRLPA-----YTSQDEKDLRLKLAFTDGEAI 1240
Qy 1019 VPQGGFCPPDAPGAGGWHHRSSSTRSGGGDLTLGLEPSEBEP-----RSPLAPS 1072
Db 1241 AKPDYLPQKAAPGPS-----HRTDCT-----DEMPKLNRYCKDPSKN 1279
Qy 1073 EGAGSDVFDG---DLGMAAGLQSLPHTDPSPLQRYSEDPVLPSETDGYVAPLTCSP 1129
Db 1280 SSTGDDERDSSAREVGVGNLR-----LDLPVEDDDYLP-TCQP 1317
Qy 1130 QPEYVNPQDVRPQPPSREGPLPAARAGATLERAKTSLPGKNGVVKDVFAGFAGVENPE 1189
Db 1318 GPNNNNMN-----NPNQNNMAVGVAGYM-----DLTGVPVSVDNPE 1356
Qy 1190 YL-----TPQGAAPQH-----PPPAFSP-AFDNLYWD 1218
Db 1357 YLLNAQTLGVGESPIQTIGIPWGGPGTMEVKVPMFPGSEPTSSDHEYND 1408

RESULT 12
ID ERBB ALV STANDARD; PRT; 634 AA.
AC P00534;
DC 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian leukosis virus.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nilssen T.W., Maroney P.A., Goodman R.G., Rottman F.M.,
RA Crittenden L.B., Raines M.A., Kung H.-J.,
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RL amino-truncated EGF receptor."
CC Cell 41:719-726 (1985).
CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -I- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -I- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M10066; AAA48763.1; ALT_INIT.
DR PIR; A00643; TVCHLV.
DR PIR; B00643; TVFVLV.
DR HSP; P11362; IFGK.
DR InterPro; IPR00719; Euk_pkinase.
DR Pfam; PF00069; pkinase.1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
DR Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
DR Glycoprotein; Phosphorylation.
DR DOMAIN 132 399 PROTEIN KINASE.
FT NP BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT SITE 257 257 BY SIMILARITY.
SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;

Query Match 24.1%; Score 1633.5; DB 1; Length 634;
Best Local Similarity 50.2%; Pred. No. 1.4e-80;
Matches 356; Conservative 79; Mismatches 141; Indels 133; Gaps 20;

Qy 587 CAHYKDPPECVARCPGVPDLSYMPFWKFPDEBAGCQPCINCHSCVDLDDKGPAAEQ 646
Db 3 CAHFDGPHCVKACPAVLGENDTL-VWKYADANAVCOLCHPNCTRGCKGPGLEGCP--- 58
Qy 647 RASPLTSIVSAVV-GILLVVLGWFGILIKRQOKIRKVTMRLLQOETELVEPLTPSGA 705
Db 59 NSKTSIAAGVVGGLLCLVVGVLGIGLYLRR-HIVRKTLLRLLQOELVEPLTPSGE 117
Qy 706 MPNQAOQMRILKETELRKVKVLGSGAFQ-----YKANSKF-----IGITELVLRNTSPKA 757
Db 118 APNQAHRLILKETEFKKVKVLGSGAFGTVVKGLWIPEGEKVKIPAIKE--LREATSPKA 175
Qy 758 NKEILDEAYVMAGVGSPPYSRLIGICLTSTVQLVTOLMPYGCILLDVRNRRGLSGODLL 817
Db 176 NKEILDEAYVMASVDNPHVCRLLIGICLTSTVQLITQMPYGCILLDVRHKNIGSQYLL 235
Qy 818 NWCQIAKMGSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLIDDETYHADGGK 877
Db 236 NWCVQIAKGMNVLDEERLVHRDLAARNVLTKPQHVKITDFGLAKLGLGADEKEYHAEGK 295
Qy 878 VPIKWMALLESILRRRTHQSDVMSYGVTVWELMTFCAKPYDGIIPAREIPDLLEKGERLPQ 937

```

```
Db 296 VPIKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGI PASBISSVLEKGERLPQ 355
Qy 938 PPICTIDVYIMVKCWMIDSCRRFRRELVSERFMRDPOFVVIQ-NEDLGPASPLDS 996
Db 356 PPICTIDVYIMVKCWMIDSCRRFRRELVSERFMRDPOFVVIQ-ODERMHLPSPDTS 415
Qy 997 TFYRSLLDDMDGLVDAEYLVPOQGFCCDPAPGAGGMVHRRHSSTRSGGDLTLG 1056
Db 416 KFYTLMEEDMEDIVDAEYLVPHQGF-----NPFST----- 449
Qy 1057 LEPSEEEAPRSL-----APSEGAGSVDFDGLGMGAAGKGLQSLTHDPSPLOQRYSDPT 1111
Db 450 -----SRTPLLSLSATSNNATNCID-----RNGQGHVPREDSPVQRYSSDPT 493
Qy 1112 VPLPSET--DGYVAPLTCPOPEYVQPDVPPQPPSPREGPLPAARFAGATLERAKTLSP 1169
Db 494 GNFLSEIDDGFL-----PAPEYVQ--LMPKXPS-----TAM 524
Qy 1170 GKNGVWQDF-----AFGGAIVENPEYLTPOGGAAPQPHPPAPSPAFDNL 1214
Db 525 VQNGIYNNISITAIKLPMSRYQNSHSTAVDNEYL-----NTNQSPLAKTVPFESS 576
Qy 1215 YYWQDQ-----DPPE-----RGAPPSTFKGTPTAENPEYLGLDVP 1248
Db 577 PYWISQGNHQINLNDPQYQDFLFPNETKPNGLLKVPAENPEYLRVAAP 625

RESULT 13
EGFR_CHICK
ID EGFR_CHICK STANDARD; PRT; 703 AA.
AC P13387;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER)
DE (Fragment).
GN EGFR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN 1;
RP SEQUENCE FROM N.A.
RX MEDLINE=88261272; PubMed=3260329;
RA Lax I., Johnson A., Hawk R., Sap J., Bellot F., Winkler M.,
RA Ullrich A., Vennstrom B., Schlessinger J., Givol D.;
RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT expression in mouse cells, and differential binding of EGF and
RT transforming growth factor alpha";
RL Mol. Cell. Biol. 8:1970-1978(1988).
CC -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M20286; AAA48760.1; -.
```

```
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; FU; 4.
DR PROSITE; PS00107; PROTEIN KINASE ATP; PARTIAL.
DR PROSITE; PS00109; PROTEIN KINASE TYR; PARTIAL.
DR PROSITE; PS00011; PROTEIN KINASE_DOM; PARTIAL.
KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1..30
FT CHAIN 31..703 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT DOMAIN 31..654 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 655..667 POTENTIAL..
FT DOMAIN 668..703 CYTOPLASMIC (POTENTIAL).
FT DISULFID 201..214 BY SIMILARITY.
FT DISULFID 222..230 BY SIMILARITY.
FT DISULFID 226..238 BY SIMILARITY.
FT DISULFID 239..247 BY SIMILARITY.
FT DISULFID 243..255 BY SIMILARITY.
FT DISULFID 258..267 BY SIMILARITY.
FT DISULFID 271..298 BY SIMILARITY.
FT DISULFID 302..314 BY SIMILARITY.
FT DISULFID 318..333 BY SIMILARITY.
FT DISULFID 336..340 BY SIMILARITY.
FT DISULFID 513..522 BY SIMILARITY.
FT DISULFID 517..530 BY SIMILARITY.
FT DISULFID 533..542 BY SIMILARITY.
FT DISULFID 546..562 BY SIMILARITY.
FT DISULFID 565..581 BY SIMILARITY.
FT DISULFID 569..589 BY SIMILARITY.
FT DISULFID 592..601 BY SIMILARITY.
FT DISULFID 605..627 BY SIMILARITY.
FT DISULFID 630..638 BY SIMILARITY.
FT DISULFID 634..646 BY SIMILARITY.
FT CARBOHYD 134..134 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT CARBOHYD 190..190 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT CARBOHYD 200..200 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT CARBOHYD 359..359 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT CARBOHYD 368..368 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT CARBOHYD 420..420 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT CARBOHYD 573..573 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT CARBOHYD 578..578 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT CARBOHYD 613..613 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT CARBOHYD 633..633 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT CARBOHYD 648..648 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT NON_TER 703..703
SQ SEQUENCE 703 AA; 77427 MW; AFP2DE11B735A690 CRC64;

Query Match 23.5%; Score 1595; DB 1; Length 703;
Best Local Similarity 44.6%; Pred. No. 1.8e-78;
Matches 316; Conservative 111; Mismatches 251; Indels 30; Gaps 14;

Qy 8 RWGLLLALLPPGAA-----STQVCTGTDMLRPLASPETHDMLRHLVYGCQVQGNLE 61
Db 13 RGAALVLLLLGVALCSAVEERKVCQGTNNKLTQLGHVEDHFTSLQRMVNNCEVLSNLE 72
Qy 62 LTYLPTNASLFLQDIQEYGVYVLIAHNOVRQVPLQRLRIVRGTLQFEDNYALAVLDNGD 121
Db 73 ITVEHNRDLTFLKTIQEVAGYVLIAMVWDVIPLENLQIIRGNVLYDSNFALAVLSNYH 132
Qy 122 PLNNTPTVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILKWDIFHNNQAL 181
Db 133 -MKNKTQ-----GLRELPMKSLSELNGVVKISNNPKLCNNMDITVLWNDIIDTSRK-PL 182
Qy 182 TLID-TNRSRACHPCSPMKGSRCEWSESSDCOSLRTTVCAGGCA-RCKGPLPTDCCHEQ 239
Db 183 TVLDFAFNSSCKPKHPCNCTEDHCWAGBQNCOTLTKVICAQCCRCRCRQVSDCCCHNQ 242
Qy 240 CAAGCTGPKHSDCLACLHFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAC 299
```

```
Db 243 CAAGCTGPRSDCLACRKRFRDADTKDTCPLPLYNPTTYQMDVNPBGKYSFGATCVREC 302
Qy 300 PYNYSLTGVSCTLVCPHNLQEVTAEDGTORCEKCKPCARVCVGLGMQVYKANSKFIGI 359
Db 303 PHNVVTDHSCVRSQNTDITYEV-EENGVRKCKCDGLCKVCNGIGIGELKGLS-INA 360
Qy 360 TELE-PAGCKKIFGLSLAFSPESFDGDPASNTAPLOPEQLQVFETLEITGYLYISAWPDS 418
Db 361 TNIDSPKNCCKINGDVSILPVAFGLDAFTKTLPLDPKLLDVFRVKEISGFLLIQAWPDN 420
Qy 419 LPDLSVFONLOVIRGRILHNGAVSLTLOGIGISWLGSLRLSRELGLALHNNHLCFVH 478
Db 421 ATDLAPENLEIRGRTKQHQGVSLAVNLKIOSLGRSLKEISDGDIAIMKNKLCYAD 480
Qy 479 TVPWDQLFRPHQALLHTANRPECEVGEGLACHQLCARGHCWGPPTQCVNCSQFRLGQ 538
Db 481 TMNWSLFAFQSKTKIQRNKNNDCTADRHVCDPLSCDVCWGPFGFHCFSRFSRQK 540
Qy 539 ECVEECVQLGRLPREYNARHCLPCHPECPQNG---SVTCFGEADQCACAHYKDPFP 595
Db 541 ECVKQCNILOGEPRERFDRSKCLPCHSECLVQNSTAYNTTCSGPGDPDKMCAHFIDGPH 600
Qy 596 CVARCPGKPDLSYMPKFWPDEGACQPCPINCSTHSCVDLDDKGCFAQRASPLTSIV 655
Db 601 CVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKPGLEGCP---NGSKTPSIA 656
Qy 656 SAVV-GILLVVLGVWFGILIKRQKIRKYTMRLLOQETELVEPLTP 702
Db 657 AGVVGGLLCVVGLGGLYLRR-HIVRKRTRLRLQLQERELVEPLTP 703

RESULT 14
ERBB AVIER
ID ERBB AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroviridae; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
gene family.";
RL Cell 35:71-78(1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84231957; PubMed=6328658;
RA Debure B., Henry C., Benaisa M., Biserte G., Claverie J.-M.,
RT Saule S., Martin P., Stehelin D.;
RT "Sequencing the erbA gene of avian erythroblastosis virus reveals a
new type of oncogene.";
RL Science 224:1456-1459(1984).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
CC -1- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
IN CHICKENS.
CC -1- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; K02006; AAA42394.1; ALT_INIT.
DR EMBL; K01216; AAA42400.1; -.
DR HSSP; A00644; TVYUH.
DR HSSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR TRANSFAC; PS00111; PROTEIN KINASE DOM; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 146 146 I -> V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MW; 76BECDD06745D609 CRC64;

Query Match 23.4%; Score 1588; DB 1; Length 604;
Best Local Similarity 50.1%; Pred.No.3.6e-78;
Matches 347; Conservative 75; Mismatches 134; Indels 136; Gaps 19;

Qy 587 CAHYKDPPEFCVARCPGKPDLSYMPKFWPDEGACQPCPINCSTHSCVDLDDKGCFAEQ 646
Db 3 CAHFIDGPHCVKACPAGVLGENDTL-VRKYADANAVCQLCHPNCTRGCKPGLEGCP--- 58
Qy 647 RASPLTSIVSNV-GILLVVLGVWFGILIKRQKIRKYTMRLLOQETELVEPLTPSGA 705
Db 59 NGSKTPSIAAGVVGGLLCVVGLGGLYLRR-HIVRKRTRLRLQLQERELVEPLTPSGE 117
Qy 706 MPNQAQMRILKETLRKRVKLVGSGAFGQ---YKANSKF---IGITELVLRNTSPKA 757
Db 118 APNQAHLRIKHETEFKVKVVLGSGAGFTYKGLWIPEGEKVKIPVAIKE--LREATSPKA 175
Qy 758 NKEILDEAYVNAVGVSPYVSRLLIGICLTSTVQLVTQIMPYGCLLDHVRENKRLGSQLDL 817
Db 176 NKEILDEAYVNAVSDNPHVCRLLIGICLTSTVQLITQIMPYGCLLDYIREHKDNIGSQYLL 235
Qy 818 NWCWQIAKMSYLEDLVRLVHRDLAARNVLKGNHVKITDFGLARLLDIDETEHADGGK 877
Db 236 NWCVQIAKGMVYLEERRLVHRDLAARNVLVKTPOHKVITDFGLAKLGADEKEYHAEGGK 295
Qy 878 VPIKWMALSIILRRRTHQSDVMSYGVTVWELMTFCAPYDGIIPAREIPDLLEKGERLPQ 937
Db 296 VPIKWMALSIILRIYTHQSDVMSYGVTVWELMTFGSGYDGIIPASEISVLEKGERLPQ 355
Qy 938 PPICTIDVYMWKVMIDSECRPRELVSFNRMDPQRFVWVQI-NEDLGPASPLDS 996
Db 356 PPICTIDVYMWKVMIDADSPKFERELIAESKWARDPPRLVLIQGDREHMLPSTDS 415
Qy 997 TFVRSLLDDDDGDLVDAEYLVPOQGFPCPDAPGAGGMVHHRHRSSTSGGGDLTLG 1056
Db 416 KFVRLTMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1057 LEPSEBEAERSPL-----APSEAGSDVFDGDLGMAAKGLQSLPHTHDSPLQYSEDP 1111
Db 450 -----SRTPLLSLSLSSATNSATNCID-----RNGQGHVPVRESFVQYRSDPT 493
Qy 1112 VPLPSET--DGVVAPLTCSPQPEYNQPDVROPSPREGPLPAARPAAGATLERAKTSLP 1169
Db 494 GNPLESIDDGL-----PAPEYVQ--LMPKKPSTAM----- 524
Qy 1170 GKNGVVKDVFAP-----GGAVENPEYLPQGGAAPQPHPPAFSPA 1210
Db 1170 GKNGVVKDVFAP-----GGAVENPEYLPQGGAAPQPHPPAFSPA 1210
```

```

118  APNOAHURLILKETFEKKVKVLGFGAFGTVYKGLWIPBEGEKVTPVPAIKELUREATSPKANK 1177
760  EILDEAYVMAGVGSPPYRSRLIGLICLTSTVOLQTLMPYGCGLLDHVRNRRGLGSQLLNW 819
178  EILDEAYVMASVDNPHVCRLLIGLICLTSTVOLITQLMPYGCGLLDYIREHKDNIGSOYLLNW 237
820  CMQIAKMSYLEDLVRLVHRDLAARNVLKSPHNVKITDFGLABLLDIDETEHADGGKVP 879
238  CVQIAKMWNYLEERHMHVRDLAARNVLKTPQHVKITDFGLAQKGADEKEYHAEGGKVP 297
880  IKWMALESILRRPETHOSDVMYSVGVTVWELMTFGAKPYDGI PARETPDLLLEKERLPOPP 939
298  IKWMALESILHRIYTHOSDVMYSVGVTVWELMTFGSKPYDGI PASEISVVLLEKERLPOPP 357
940  ICTIDVMIWMKCMIDSECRPFRELVLVSFEFSWARDPQRFVVIQ--NEDLGPASPLDSTF 998
358  ICTIDVMIWMKCMWDSDSRPFRELILABFSKWARDPPRYLVIQDERMHLPSPTDSKF 417
999  YRSILLEDDDDGLVDABEYLVPOGGFFCPDPAPFAGGVMVHRRHSRSTSGGGDLTLGLE 1058
418  YRTLWEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
1059  PSEBEADPSPL-----APSGAGSDVPDGLGMGAAKGLQSLTHDPSPLOYSEDPSTVP 1113
450  -----SRTPLLSLSLSTSNSTNCIDRNG-----H----- 476
1114  LPSETDGVAPLTCSPQPEYVNPQDVRPOPPSPREGPLPAARPAGAT--LERAKTLSPGKN 1172
477  -PVREDGFL-----PAPEYVNO--LMPKKPSTAMVQNIYVILSTAIKSLPDSRYQN 527
1173  GVVKDVPFAGGAVENPEYL 1191
528  -----SHSTAVDNPEYL 539

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:01:42 ; Search time 46.6008 Seconds
(without alignments)
5522.503 Million cell updates/sec

Title: SEQ4-710-730-12
Perfect score: 6775
Sequence: 1 MEALALCRMGILLALLPPGA.....TFKGTPTAENPEYGLDVPV 1249

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPREMBL 21:*
- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6091	89.9	1259	6 O18735	O18735 canis famil
2	3028	44.7	1209	11 Q9QX70	Q9QX70 rattus norv
3	2999	44.3	1210	11 Q9EP98	Q9EP98 mus musculu
4	2611	38.5	1165	13 Q9YH40	Q9YH40 xiphophorus
5	2577.5	38.0	1137	13 Q9W6F6	Q9W6F6 gallus gall
6	2191	32.3	1328	13 P79754	P79754 fugu rubrip
7	1930.5	28.5	1433	5 Q9BIH9	Q9BIH9 anopheles g
8	1871	27.6	419	4 Q9UK79	Q9UK79 homo sapien
9	1739	25.7	367	11 Q8RX21	Q8RX21 mus musculu
10	1697.5	25.1	412	4 Q8WYV0	Q8WYV0 homo sapien
11	1604	23.7	729	15 Q86712	Q86712 avian rous-
12	1602	23.6	567	15 Q86714	Q86714 avian rous-
13	1543.5	22.8	962	15 Q64895	Q64895 avian eryth
14	1535	22.7	545	15 Q85468	Q85468 avian eryth
15	1506.5	22.2	655	11 Q9WVF5	Q9WVF5 mus musculu
16	1490.5	22.0	643	11 Q9ERV6	Q9ERV6 mus musculu

17	1206	17.8	1193	5 Q9YIX8	Q9YIX8 ephydatia f
18	1126	16.6	527	13 Q90836	Q90836 gallus gall
19	1125.5	16.6	1368	5 Q23821	Q23821 caenorhabdi
20	1109	16.4	1717	5 Q26566	Q26566 schistosoma
21	1001.5	14.8	478	11 Q9ESE0	Q9ESE0 rattus norv
22	942.5	13.9	599	13 Q9PSH2	Q9PSH2 gallus gall
23	906	13.4	165	4 Q14256	Q14256 homo sapien
24	887	13.1	176	11 Q923V5	Q923V5 rattus norv
25	806.5	11.9	346	13 P11776	P11776 xiphophorus
26	778	11.5	435	5 Q8SZW1	Q8SZW1 drosophila
27	754.5	11.1	311	13 Q99162	Q99162 xiphophorus
28	734	10.8	331	4 Q9BUD7	Q9BUD7 homo sapien
29	723	10.7	149	6 Q9B666	Q9B666 oryctolagus
30	698.5	10.3	1362	13 Q9PVZ4	Q9PVZ4 xenopus lae
31	682	10.1	1671	5 Q9NJV5	Q9NJV5 biophalar
32	667.5	9.9	1368	13 Q8UW85	Q8UW85 paralicth
33	659	9.7	1418	13 Q93457	Q93457 scophthalmu
34	640.5	9.5	1369	13 Q8UW86	Q8UW86 paralicth
35	634.5	9.4	1472	5 Q9U5A8	Q9U5A8 bombyx mori
36	625	9.2	1412	13 Q8UW84	Q8UW84 paralicth
37	620	9.2	1358	13 Q73798	Q73798 xenopus lae
38	608.5	9.0	1418	13 Q8UW83	Q8UW83 paralicth
39	601	8.9	1245	13 Q9YGH8	Q9YGH8 scophthalmu
40	588	8.7	1371	11 Q9QVW4	Q9QVW4 rattus sp.
41	581.5	8.6	1091	4 Q9UMQ4	Q9UMQ4 homo sapien
42	579	8.5	987	11 Q91YMO	Q91YMO mus musculu
43	578.5	8.5	2144	5 Q9VD94	Q9VD94 drosophila
44	576	8.5	935	4 Q96L35	Q96L35 homo sapien
45	576	8.5	987	11 Q99MR2	Q99MR2 mus musculu

ALIGNMENTS

RESULT 1

O18735	ID	O18735	PRELIMINARY;	PRT;	1259 AA.
AC	O18735;				
DT	01-JAN-1998	(TrEMBLrel. 05, Created)			
DT	01-JAN-1998	(TrEMBLrel. 05, Last sequence update)			
DT	01-JUN-2002	(TrEMBLrel. 21, Last annotation update)			
DE	ErbbB-2.				
OS	Canis familiaris (Dog).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.				
OX	NCBI_TaxID=9615;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Yokota H.;				
RT	"CDNA cloning of erbB-2 from canine mammary gland.";				
RL	Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AB008451; BAA23127.1; -				
DR	HSSP; P11362; 1FGK.				
DR	InterPro; IPR002048; EF-hand.				
DR	InterPro; IPR000494; EGFR_L_domain.				
DR	InterPro; IPR000719; Euk_pkinase.				
DR	InterPro; IPR001245; Tyr_pkinase.				
DR	InterPro; IPR004019; YLP_motif.				
DR	Pfam; PF00757; Furin-like; 1.				
DR	Pfam; PF00069; pkinase; 1.				
DR	Pfam; PF01030; Recep_L_domain; 2.				
DR	Pfam; PF02757; YLP_2.				
DR	ProDom; PD000001; Euk_pkinase; 1.				
DR	SMART; SM00261; FY; 3.				
DR	SMART; SM00219; TyrcK; 1.				
DR	PROSITE; PS00018; EF_HAND; UNKNOWN 1.				
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.				
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.				
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.				
KW	ATP-binding; Transferase; Tyrosine-protein kinase.				
SQ	SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;				

```
Query Match      89.9%; Score 6091; DB 6; Length 1259;
Best Local Similarity 89.5%; Pred. No. 0;
Matches 1128; Conservative 44; Mismatches 75; Indels 14; Gaps 4;

Qy 1 MELAAALCRWGLLLALLPSPGAASCTGCTDKMLRLPASPTHLDMLRHLVGGCGVVGNNL 60
Db 1 MELAAWCRWGLLLALLPSPGAAGTCTGCTDKMLRLPASPTHLDMLRHLVGGCGVVGNNL 60

Qy 61 ELTVLPNASSLSFLQDIQEVQGVYLIHAHNOVROVPLQRLRIVRGTQLPEDNYALAVLDNG 120
Db 61 ELTVLPANASSLSFLQDIQEVQGVYLIHAHNOVROVPLQRLRIVRGTQLPEDNYALAVLDNG 120

Qy 121 DPLNNTTVPVTGASPGGLRELRLQLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNQOLA 180
Db 121 DPLEGGIPAPGAAQGLRELRLQLSLTEILKGGVLIQRNPOLCHQDTILWKDVHKNQOLA 180

Qy 181 LTLIDTNRSRACHPCSPMCKGSRGWSSEDCOSLTRVTCAGGCARCKGPLPTDCCHQEC 240
Db 181 LTLIDTNRFRSACPPSPACKDAHCWAGSSGDCOSLTRVTCAGGCARCKGQPTDCCHQEC 240

Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTSCP 300

Qy 301 YNYLSTDVGSCTLVCPLNHOEVTAEDGTQRCCKSKPCARVCYGLGMQYIKANSKFTGIT 360
Db 301 YNYLSTDVGSCTLVCPLNHOEVTAEDGTQRCCKSKPCARVCYGLGNEHLREVRAVTSAN 360

Qy 361 ELEPAGCKKIFGSLAFIPESFDGDPASNTAPLOPEQLQVFETLEETGYLYISAWPDSLP 420
Db 361 IQEPAGCKKIFGSLAFIPESFDGDPASNTAPLOPEQLRVFEALEETGYLYISAWPDSLP 420

Qy 421 DLSVFQNLQVTRGILHNGAYSILTLOGIGISWGLRSLRGLSGSLALIHNTLFCVHTV 480
Db 421 NLSVFQNLVIRGVLDHNGAYSILTLOGIGISWGLRSLRGLSGSLALIHNRALCFVHTV 480

Qy 481 PWDOLFNPQHALLHTANRPEDECEVCGELACHQICARGHCWGPGPTQVCNCSQFLRGQEC 540
Db 481 PWDOLFNPQHALLHSANRPEECVCGELACYP-CAHGHCKWGPGPTQVCNCSQFLRGQEC 539

Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPEADQCACAHYKDPFPCVARC 600
Db 540 VEECRVLQGLPREYVNDRYCLPCHSECQPNGSVTCFSGEADQCACAHYKDPFPCVARC 599

Qy 601 PSGVKPDLNMYPTWKPDESGACQPCPINCTHSCVDLDDKGCBAEQRASPLTSTVSAVG 660
Db 600 PSGVKPDLNMYPTWKPDESGACQPCPINCTHSCVDLDDKGCBAEQRASPLTSTVSAAVG 659

Qy 661 ILLVVLGVVFGIILIKRRQKIRKYTMRLLIQETELVEPLTPSGAMPNQAMRILKETEL 720
Db 660 ILLAVVGLVLGILIKKRRQKIRKYTMRLLIQETELVEPLTPSGAMPNQAMRILKETEL 719

Qy 721 RKVKVLGSGAFGVYKA-----NSKFIGITELVLRENTSPKANKEILDEAYVMAGVGS 773
Db 720 RKVKVLGSGAFGVYKGIWIPDGENVR-IPVAIKVLRENTSPKANKEILDEAYVMAGVGS 778

Qy 774 PYVSRLAGICLTSTVQLVLTQMPVGCGLDHHVRENRGLSGQDLLNWCQIAKGMVSLDY 833
Db 779 PYVSRLAGICLTSTVQLVLTQMPVGCGLDHHVRENRGLSGQDLLNWCQIAKGMVSLDY 838

Qy 834 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHYADGGKVPKKNMALESILRRRF 893
Db 839 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHYADGGKVPKKNMALESIPRRF 898

Qy 894 THOSDVMWSYGVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMWVKCW 953
Db 899 THOSDVMWSYGVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMWVKCW 958

Qy 954 MIDSECPRELVSEFSRWARPOREVFVIONEDLGASPLDSTFYSLLEDDMDGLVD 1013
Db 959 MIDSECPRELVSEFSRWARPOREVFVIONEDLGASPLDSTFYSLLEDDMDGLVD 1018

Qy 1014 AEEYLVPPQGFPCPTGAGGTAAHRRHSSSTRSGGDLTGLGPESEBPPKSLAPSE 1073
Db 1019 AEEYLVPPQGFPCPTGAGGTAAHRRHSSSTRSGGDLTGLGPESEBPPKSLAPSE 1078

Qy 1074 GAGSDVDFDGLNGAAKGLQSLPDRYSEDPTVPLPSETDGYVAPLTCSPQPEY 1133
Db 1079 GAGSDVDFDGLNGAAKGLQSLPDRYSEDPTVPLPSETDGYVAPLTCSPQPEY 1138

Qy 1134 VNQPDVVRPQPPSPREGPLPAARPAAGATLER-----AKTILSPGKNGVVKVDFATGGAVENP 1188
Db 1139 VNQPEVWPQPLALEGPLPPSRPAGATLERPKTILSPKTLSPGKNGVVKVDFATGGAVENP 1198

Qy 1189 EYLTPQGGAAPOPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPBYLGLDVP 1248
Db 1199 EYLAPRGAAPQPHPPAFSPAFDNLVYWDQDPPSRGSPSTFEGTPTAENPBYLGLDVP 1258

Qy 1249 V 1249
Db 1259 V 1259

RESULT 2
Q9QX70 PRELIMINARY; PRT; 1209 AA.
AC Q9QX70;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor.
GN EGFR.
OS Rattus norvegicus (Rat).
OC Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RX MEDLINE=90258888; PubMed=2342466;
RA Petch L.A., Harris J., Raymond V.W., Blaeband A.J., Lee D.C.,
RA Earp H.S.;
RT "A truncated, secreted form of the epidermal growth factor receptor is
RT encoded by an alternatively spliced transcript in normal rat tissue."
RL Mol. Cell. Biol. 10:2973-2982(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RA Petch L.A.;
RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RA Guttridge K., Dawson T.L., Earp H.S.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; M37394; AAF14008.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
KW ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1209 AA; 134891 MW; 96FEE7F6CC1B7773 CRC64;

Query Match      44.7%; Score 3028; DB 11; Length 1209;
Best Local Similarity 49.1%; Pred. No. 8.6e-221;
```


Matches 629; Conservative 165; Mismatches 360; Indels 126; Gaps 31;

```
QY 3 LAALCRWGLLLALPPGA-ASTQVCTGTDMKRLPASPETHLDMRLHYQCGQVQCNLE 61
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
15 LAALCAAG-----GALBEKKVCCQGSNRLTQGLTFEDHFLSLQRMFNNEVVLGNLE 66
QY 62 LTYLPTNASLFLDIOIEGVYLIJAHQVQVPLQRLIRVGTQQLFEDNYALAVLNDG 121
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
67 ITVQRNYDLSFLKTIQEVAGYVLIJAHQVQVPLQRLIRVGTQQLFEDNYALAVLNDG 124
QY 122 PLNNTTPTVTCASPGGLRELOLRSITLTKGVLVLRNPOLCYODTILWKDIFHKNQ 181
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
125 -----YGTNTKGLRELPMRNLQELIGAVRFSNNPILCNMETIQWRDIV-ODV 175
QY 182 TLIDTNRS-RACHPCSPMCKGRSGWGSSEDCQSLTRTVCAAGCA-RCKGPLPTDC 239
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
176 MSMDVQRHLTGCPKCDPSCPNCGWGRGEENCQKLTIIICAQCCSRRCGRSPSDCC 235
QY 240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMNPBGRYTFGAS 299
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
236 CAAGCTGPRSDCLVCHRFDEATCKDTCPLMLYNTTYQMDVNPBGKYSFGATCVK 295
QY 300 PYNVLSLDVSGCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMQVIKANS 359
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
296 PRNVFVTDHSGVACAGPDYIEV-EEDGVSKCKKDCGPKRVCNIGIGIBFK-DT 353
QY 360 TELE-FAGCKKIFGSLAFSPESFDGDPASNTAPLQPEQLQVFTLEBITGYLVI 418
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
354 TNIKFKYCTAISGLHLIPVAFKDSFTPTPLDPRELEILKTVKEITGFLLIQAW 413
QY 419 LPDLSVQNIQVIRGRILHNGAVSLTQGLGISWGLRSRLGSLGIALIHNTHLC 478
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
414 WTDLHAPENLEIRGRTKQHQGFSLAVVGLNITSLGRLSLKEISDGDVITSGNR 473
QY 479 TVPQDQLFRPHQALLHTANRDECEVGEGLACHOLCARGHCWGPCTQVNCSPQL 538
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
474 TINWKLFGFNPQKTKIMNRAEKDKCATNHVCNPLCSSEGCWGPETDCVSCQNV 533
QY 539 ECVEECRVLOGLPREYVYNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDP 598
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
534 ECVDKCNILGEPRFVENSECLOCHPECLPQTNITCTGRGPNDCIKAHYVDPCH 593
QY 599 RCPGSGVKPDLISYMPIWKPPEEGACQPCPINCTHSCVDLDDKCPAFQORASP 657
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
594 TCSPGINGENNTL-VWKFADANNVCHLCHANCTYGCAGPGLKGC--QQPEGPKI 650
QY 658 VUGLLVVLGVVFGI-LIKRQOKIRKYTWRLLOSTELVEPLTPSGAMPNQAQRIL 716
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
651 IVGGLLFIVV-VALGIGLFWRRQLVRKRTLRLRLQERELVEPLTPSGEAPNQA 709
QY 717 ETELKRVKVLGSGAFGQ-----YIKANSKF---IGITELVRENTSPKANKEIL 768
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
710 ETEFKKIVLGSAGFVYKGLWIPGEKVKIPVAIKE--LREATSPKANKEILDEAY 767
QY 769 AGVGSPPVSRLLGICLSTVQLVTPQMPYCGCLLDHVRNRRGLRQSDLLNCWQIA 828
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
768 ASVDNPHVCRLLGICLSTVQLITQMPYCGCLLDYVREHKDNIGSYQLLNCWQIA 827
QY 829 YLEDVRLVHRDLAARNVLKSPNHVKITDGLARLLDIDETEYHADCGKVPKMALE 888
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
828 YLEDVRLVHRDLAARNVLKSPNHVKITDGLARLLDIDETEYHADCGKVPKMALE 887
QY 889 LRRRFTHQSDVWSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPIC 948
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
888 LHRITYHQSDVWSYGVTVWELMTFGSKPYDGPAREIPDLLEKGERLPQPPIC 947
QY 949 MVKCMWIDSECRPRFRELVEFSRMDRQRFVVIQ-NEDLGPASPLDSTFYRSL 1007
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
948 MVKCMWIDSECRPRFRELVEFSRMDRQRFVVIQ-NEDLGPASPLDSTFYRSL 1007
QY 1008 MGDVLDAEELVLPQCGFCDDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEA 1067
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1008 MEDVVDADAEYLIPQCGFF-----NSPST-----SRT 1033
```

```
QY 1068 PLAPSEGAGSDVFDGLGMGAAGKGLQSLPHTDPSPLQRYSEDDTVPPLPSET-- 1125
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1034 PLLSSLSANSN-----SSTVACINNRNGSCRKVEDAFQRYSSDPTSVLTEDNID 1086
QY 1126 TCSPOBEYVNOPDVRRPQSPREGPLPAARPAGATLERAKTLPKNGVGVKDVFA 1185
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1087 ---PVPEYINQ-SVPRPAGSVQNPVYHNQPLHP-----APGRDLHYQN--PH 1132
QY 1186 ENPEYL-TQGGGAAPQHPHPPAFSPADNLYYWDQ-----DP-----PER 1228
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1133 SNPEYLNIAQ-----PTCLSSGFDSSALWIKGSHOMSLDNDYQDFFPKAKPN 1183
QY 1229 STFPGTPTAENPEYLGLDVDP 1248
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1184 GIFKG-PTAENAEYLRVAPP 1202

RESULT 3
Q9EP98 PRELIMINARY; PRT; 1210 AA.
AC Q9EP98;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor isoform 1.
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTA;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Mailhe N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Schehl C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Mailhe N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF275366; AAG28045.1; -.
DR EMBL; AF275364; AAG28045.1; JOINED.
DR EMBL; AF275365; AAG28045.1; JOINED.
DR EMBL; AF275367; AAG24386.1; -.
DR HSSP; P11362; 1FGK.
DR MGD; MGI:95294; Egfr.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR PRODOM; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00220; S_TK; 1.
DR SMART; SM00219; TyrKc; 1.
```



```
Qy 489 PHQALLHNPANPEDECVGEGLACHQLCARGHCWGPGPTQCVCNCSQFLRGQECVBECEVLQ 548
Db 450 SDHEVMVQKRNATECHEEGMECEQSCAGCAGKQPEQCLCKNVKYKGLDSCK--- 506
Qy 549 GLPREY-VNARHCLPCHPECPQNGSVTCFPEADQCVACHYKDPFCVARCP----- 601
Db 507 SLPLRYSDSKTGDCHQCKD-----FCYPNEDNCSCNMVNDGRFCVAECPTTKHAM 561
Qy 602 -----SGVKPDLSPYPIWKRPD----- 618
Db 562 NGTCINCHKTCVCGRPRDTIAPGCTICSDKAIIGSDAKIERCLMKDESCPDGYSDYVL 621
Qy 619 -BEG----- 621
Db 622 QEEGLQKLSKAVCRKCHPRCKKCTGYGFHEQFCQECTGYKGEQCEDECPQDFYANEE 681
Qy 622 --ACQCPINCT-----HSCVDL-----DD-----KGCPAEQ----- 646
Db 682 TRICLPCHQECRGCHGLGDDHHECRNLKLFEGDPYDNATTTFCVSNCPASHPYKRFPOEA 741
Qy 647 -----RASPLTSIVSAVVGILLVVVLGVVFGI---LIKRRQKIRKYTM 687
Db 742 GKIGPYCSADSMQGLRIEPTQVKIVGMSVMAILLICVFGIAFVLSRHNKKDAVKM 801
Qy 688 RRLQETELVEPLTFPGAMPNQAMQRIKTELKRVKVLGSGFARQYIK-----ANSKF 741
Db 802 TMLAGCEDSPLRPSNVGNLTKLRIKEAEIRRGVGLMGAFGRVFKGVMPEGESVK 861
Qy 742 IGITELVIRENTSPANKELIDELAYVAGVSPVSRLLGLCLTSTVOLTMPLPYCGLL 801
Db 862 IPVAIKVLMEMSGSESKFELEAYIMASVEHPNLLKLLAVCMTSQOMLITQLMPLGCLL 921
Qy 802 DHVRENGRLGSODLLNWCIOAGMSVLEDRVLVHRDLAARNLVKSPNHVKITTDGLA 861
Db 922 DYVRNKKDKISKALLNWSIQIARGMAYLEBRRVLVHRDLAARNLVQTPSCVKITVGLA 981
Qy 862 RLLDIDETEHADGGKVPKIMMALESILRRFTHCSDVMSYGVTVWELMTFGAKPYDGIP 921
Db 982 KLLDFDSEYRAAGKMPIKWLALCEIRHRVFTSKSDVWAFGITIWELLITGARPENVP 1041
Qy 922 AREIPDLLEKERLPPOPICTIDYIMVMKWMIDSECRPRPRELVSEFSFEMARDPORFV 981
Db 1042 AKOYPELIEIGHKLPOPDICSLDYCYILLSCWILDADARPTFKLAETFAEKARDPGRYL 1101
Qy 982 VIQNEGLGASPLDSTFVRSLEDDMDGLV----- 1012
Db 1102 MI-----PGDKFRLPSYTNQDEKDLIRTLAPVMAAAAAAAGASNVDPVPTIA 1152
Qy 1013 DAEYLVPOQGFPCDPAPAGAGMVRHRRSSSTRSGGDLITLGLFSESEAPRS----- 1067
Db 1153 ETDEYLQPKTRPSIMLPQPSA-----VEPS-DEMPKSLRYCK 1188
Qy 1068 -PLAP---SEGAGSDVFDGLGMAAGKGLQSLPHDPSPLORYSEDTVPPLPSTDGYVA 1123
Db 1189 DPLKPDDETGDHGEV-----GVGGIR-----LNLPLDEDDYLM 1222
Qy 1124 PLTCSPOPEYVQDVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKDVFAGG 1183
Db 1223 P-TCQSQ-----NQS-----TPG-----YMDLIGVPA 1243
Qy 1184 AVENPEYL-----TPQGAAPQPHPPAFSPAFDNLVYWDQDPPERGAPSTFKGT 1234
Db 1244 SVDNPEYLMGSTOAILAGLQSGMG--PHTPP-----PNTFNGM 1280
Qy 1235 PTAENPE 1241
Db 1281 PTHQHSQ 1287

RESULT 8
Q9UK79
ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
```

```
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF177761; AAD56009.2; -.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 1.
DR SMART; SMO0261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FEC1BE347E2D030C CRC64;

Query Match 27.6%; Score 1871; DB 4; Length 419;
Best Local Similarity 98.8%; Pred. No. 1.3e-133;
Matches 341; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MELAAALCRWGLLALLPPGAASQVCTGTDMLRLPASPTHLDMLRLHLYQGCVVQGNL 60
Db 1 MELAAALCRWGLLALLPPGAASQVCTGTDMLRLPASPTHLDMLRLHLYQGCVVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQGVVLIAHNQVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQGVVLIAHNQVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNOLA 180
Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNOLA 180
Qy 181 LTLIDNRSRACHPCSPMKSGSCWGESSEDCSLTRTVCAGGCARCKGLPTDCCHEQC 240
Db 181 LTLIDNRSRACHPCSPMKSGSCWGESSEDCSLTRTVCAGGCARCKGLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCLPHNQEVTAEDGTQRCCKSPCARVCYGL 345
Db 301 YNYLSTDVGSCTLVCLPHNQEVTAEDGTQRCCKSPCARVTHSL 345

RESULT 9
Q8R2X1
ID Q8R2X1 PRELIMINARY; PRT; 367 AA.
AC Q8R2X1;
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
```

DR	ENBL; BC027080.1; --
KW	Hypothetical protein.
SQ	SEQUENCE 367 AA; 40163 MW; OBE03395F9E101B0 CRC64;

Query Match	
Best Local Similarity	25.7%; Score 1739; DB 11; Length 367;
Mismatches	88.0%; Pred. No. 1.1e-123;
Matches	323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

QY	883 MALESILRRRFTHQSDVMSYGVTVMELMTFGAKPYDGIPIAREIPDLLEKGERLPPQPCT 942
DB	1 MALESILRRRFTHQSDVMSYGVTVMELMTFGAKPYDGIPIAREIPDLLEKGERLPPQPCT 60
QY	943 IDVYMIMVKCWMIDSECRPRRELVSERFMRARDPQRVFIQNEIDLGPASPLDSTFYRSL 1002
DB	61 IDVYMIMVKCWMIDSECRPRRELVSERFMRARDPQRVFIQNEIDLGPASPLDSTFYRSL 120
QY	1003 LEDDDMGDLVADEEYLVPQQGFCDPAPGAGGVHHRSSSTRSGGGDLTLGLEPSEE 1062
DB	121 LEDDDMGDLVADEEYLVPQQGFCDPAPGAGGVHHRSSSTRSGGGDLTLGLEPSEE 180
QY	1063 EAPRSP LAPSEGAGSDVF DGLGMGAAGLQSLTPHDPSPLQRYSEDPTVLPSETDGYV 1122
DB	181 EAPRSP LAPSEGAGSDVF DGLGMGAAGLQSLTPHDPSPLQRYSEDPTVLPSETDGYV 240
QY	1123 APLTCSPOPEYNQDVVRPQSPREGPLPAARPAAGATLERAKT LSPGKNVGVDVFAFG 1182
DB	241 APLTCSPOPEYNQDVVRPQSPREGPLPAARPAAGATLERAKT LSPGKNVGVDVFAFG 300
QY	1183 GAVENPEYLTPOGGAAPHPAPPASPFDNLYYWDOPPERGAPSTFKGTPTAENPEY 1242
DB	301 GAVENPEYLA PRAGTAGTASQHPSPAPFDNLYYWDQNSSEQGPPPFTEGTPTAENPEY 360
QY	1243 LGLDVFPV 1249
DB	361 LGLDVFPV 367

RESULT 10	
ID	QBWYVO PRELIMINARY; PRT; 412 AA.
AC	QBWYVO
DT	01-MAR-2002 (TrEMBLrel. 20, Created)
DT	01-JUN-2002 (TrEMBLrel. 20, Last sequence update)
DT	01-MAR-2002 (TrEMBLrel. 21, Last annotation update)
DE	Hypothetical 44.7 kDa protein.
GN	P31659.
OS	Homo sapiens (Human)
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T., Wan D.F., Gu J.R.;
RT	"Novel human cDNA clones with function of inhibiting cancer cell growth."
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR	ENBL; AF318349; AAL55956.1; --
DR	InterPro; IPR002048; EF-hand.
DR	InterPro; IPR000719; Euk_pkinase.
DR	InterPro; IPR001245; Tyr_pkinase.
DR	pfam; PF00069; pkinase; 1.
DR	pfam; PF02757; VLP; 2.
DR	ProDom; PD000001; Euk_pkinase; 1.
DR	SMART; SM00219; Tyrcg; 1.
DR	PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR	PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW	Hypothetical protein.
SQ	SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match	
Best Local Similarity	25.1%; Score 1697.5; DB 4; Length 412;
Mismatches	80.5%; Pred. No. 1.9e-120;

QY	883 MALESILRRRFTHQSDVMSYGVTVMELMTFGAKPYDGIPIAREIPDLLEKGERLPPQPCT 942
DB	1 MALESILRRRFTHQSDVMSYGVTVMELMTFGAKPYDGIPIAREIPDLLEKGERLPPQPCT 60
QY	943 IDVYMIMVKCWMIDSECRPRRELVSERFMRARDPQRVFIQNEIDLGPASPLDSTFYRSL 1002
DB	61 IDVYMIMVKCWMIDSECRPRRELVSERFMRARDPQRVFIQNEIDLGPASPLDSTFYRSL 120
QY	1003 LEDDDMGDLVADEEYLVPQQGFCDPAPGAGGVHHRSSSTRSGGGDLTLGLEPSEE 1062
DB	121 LEDDDMGDLVADEEYLVPQQGFCDPAPGAGGVHHRSSSTRSGGGDLTLGLEPSEE 180
QY	1063 EAPRSP LAPSEGAGSDVF DGLGMGAAGLQSLTPHDPSPLQRYSEDPTVLPSETDGYV 1122
DB	181 EAPRSP LAPSEGAGSDVF DGLGMGAAGLQSLTPHDPSPLQRYSEDPTVLPSETDGYV 240
QY	1123 APLTCSPOPEYNQDVVRPQSPREGPLPAARPAAGATLERAKT LSPGKNVGVDVFAFG 1182
DB	241 APLTCSPOPEYNQDVVRPQSPREGPLPAARPAAGATLERAKT LSPGKNVGVDVFAFG 300
QY	1183 GAVENPEYLTPOGGAAPHPAPPASPFDNLYYWDOPPERGAPSTFKGTPTAENPEY 1242
DB	301 GAVENPEYLA PRAGTAGTASQHPSPAPFDNLYYWDQNSSEQGPPPFTEGTPTAENPEY 360
QY	1243 LGLDVFPV 1249
DB	361 LGLDVFPV 367

RESULT 11	
ID	QB6712 PRELIMINARY; PRT; 729 AA.
AC	QB6712
DT	01-NOV-1996 (TrEMBLrel. 01, Created)
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE	Polyprotein.
GN	POLYPROTEIN
OS	Avian rous-associated virus type 1.
OC	Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX	NCBI_TaxID=11950;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Medline=94203659; PubMed=8152791;
RA	Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V., Johnson A., Beug H.;
RT	"Retrosiral capture of c-erbB proto-oncogene sequences: rapid evolution of distinct viral genomes carrying mutant v-erbB genes with different transforming capacities.";
RL	Oncogene 9.1307-1320(1994).
DR	ENBL; S69372; AAC60725.1; --
DR	HSP; P03322; IAGS.
DR	InterPro; IPR000719; Euk_pkinase.
DR	InterPro; IPR004028; Retro_M.
DR	InterPro; IPR001245; Tyr_pkinase.
DR	pfam; PF00069; pkinase; 1.
DR	pfam; PF02813; Retro_M; 1.
DR	ProDom; PD000001; Euk_pkinase; 1.
DR	SMART; SM00219; Tyrcg; 1.
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR	PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW	ATP-binding; transferase; tyrosine-protein kinase.
SQ	SEQUENCE 729 AA; 80649 MW; 84D2F6914EFFE1D63 CRC64;

Query Match	
Best Local Similarity	23.7%; Score 1604; DB 15; Length 729;
Mismatches	52.5%; Pred. No. 5.5e-113;
Matches	344; Conservative 75; Mismatches 124; Indels 112; Gaps 18;

QY	569 PQNGSVTCFPGADQCACAHYKDPFCVARCPGVKPDLSYMPFWKPBEGACQPCPI 628
----	--

```
Db 141 PEETATPKTGP--DHCMCAHFIDGPHCVKACPAAGVLGENDTL-VMKYADANAVCQLCHP 197
Qy 629 NCHTSCVDLDKGCAPAEQASPLTSIVSAV-GILLVVVLGVVFGILIKRQOKIRKYTM 687
Db 198 NCTGCKGPGLEGCP--NGSKTSPAGVGGGLCLVVGGLGIGLYLRRR-HIVRKRTL 253
Qy 688 RRLQETELVPLTPSGAMPNQAORILKETELRKVKVLGSGAFGQ-----YIKANSKF- 741
Db 254 RRLQERELVPLTPSGEAPNQAHLRIKETEFKKVKVLGSGAFGTVYKGLWIPGEKVK 313
Qy 742 --IGITELVRENTSPKANKEILDEAYVMAGVSPYVSRLLGICLTSTVQLVTOLMPYGC 799
Db 314 IPVAIKE--LREATSPKANKEILDEAYVMASVDNPRVCRLLGICLTSTVQLITQMPYGC 371
Qy 800 LLDHVRENRGLSGODLLNWCQIAKMSYLEDLVLRDLAARNVLKSPNHVKITDFG 859
Db 372 LLDVIREKHONIGSOYLLNWCQIAKGMNYLEERLVRDLAARNVLKTPQHVKITDFG 431
Qy 860 LARLLDIDETEHADGGKVPKMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDG 919
Db 432 LAKLGADEKEYHAEGGKVPKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDG 491
Qy 920 IPAREIDPLEKGRLLPQPPCTIDVYIMVKMWIDSECRPRPRELVSFBSRMARDPQR 979
Db 492 IPASEISSVLEKGRLLPQPPCTIDVYIMVKMWIDADSRPKFRELIAEFSKWARDPPR 551
Qy 980 FVLIQ-NEDLGPASPLDSTFYSRLLEDMDGLVDAEYLVPOGFCFDPAPGAGGMVH 1038
Db 552 YLVIQGERMHLPSPTDSKFYRTLMEBEDMEDI VDAEYLVPHQGF----- 598
Qy 1039 HRHRSSTRSGGDLTLGLEPSEEAAPRSP-----APSEGAGSDVDFDGLMGAAKGLQ 1093
Db 599 ---NSPST-----SRTPLLSLSATSNSATNCID-----RNGQ 629
Qy 1094 SLPHDPSLPQRYSEDPTVPLPSET--DGYVAPLTCSPQPEYVNPQDVRPQPPSPREGPL 1151
Db 630 GHPVREDSFVQRYSSDPTGNFLEESIDDGFL-----PAPEYVNO--LMPKKPS----- 675
Qy 1152 PAARPAGATLERAKTLPFGKGVKVDVF-----AFGGAVENPEYL 1191
Db 676 -----TAMVQNIYNNISLTAISKLPMDSRVQNSHSTAVDNPEYL 715

RESULT 12
Q6714
ID Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; -.
DR HSP; P11362; IFGK.
DR InterPro; IPR00719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase.1
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrcK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
```

```
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Tyrosine-protein kinase.
FT NON_TER 1
SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725E1 CRC64;

Query Match 23.6%; Score 1602; DB 15; Length 567;
Best Local Similarity 53.1%; Pred. No. 5.3e-113;
Matches 343; Conservative 73; Mismatches 118; Indels 112; Gaps 18;

Qy 578 GPEADQCVACAHYKDPFPCVRCPSGVKPDLSYMPYIWKPFDEGACQCPCEINTHSCVDL 637
Db 1 GP--DHCMCAHFIDGPHCVKACPAAGVLGENDTL-VMKYADANAVCQLCHPNCRTGCKGP 57
Qy 638 DDKGCAPAEQASPLTSIVSAV-GILLVVVLGVVFGILIKRQOKIRKYTMRLLOETEL 696
Db 58 GLEGCP--NGSKTSPAGVGGGLCLVVGGLGIGLYLRRR-HIVRKRTLRLLOEREL 113
Qy 697 VEPLTPSGAMPNQAORILKETELRKVKVLGSGAFGQ-----YIKANSKF--IGITELV 748
Db 114 VEPLTPSGEAPNQAHLRIKETEFKKVKVLGSGAFGTVYKGLWIPGEKVKIPVAIKE-- 171
Qy 749 LRENTSPKANKEILDEAYVMAGVSPYVSRLLGICLTSTVQLVTOLMPYGCCLLDHVRENR 808
Db 172 LREATSPKANKEILDEAYVMASVDNPRVCRLLGICLTSTVQLITQMPYGCCLLDYIREHK 231
Qy 809 GRIGSQDLLNWCQIAKMSYLEDLVLRDLAARNVLKSPNHVKITDFGLARLLDIDE 868
Db 232 DNTGSQYLLNWCQIAKGMNYLEERLVRDLAARNVLKTPQHVKITDFGLAKLGADE 291
Qy 869 TEYHADGGKVPKMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPARIPDL 928
Db 292 KEYHAGGKVPKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPASEISSV 351
Qy 929 LEKGERLPQPPCTIDVYIMVKMWIDSECRPRPRELVSFBSRMARDPQRFVVIQ-NED 987
Db 352 LEKGERLPQPPCTIDVYIMVKMWIDADSRPKFRELIAEFSKWARDPPRYLVIQGER 411
Qy 988 LGPASPLDSTFYSRLLEDMDGLVDAEYLVPOGFCFDPAPGAGGMVHHRRSSSTR 1047
Db 412 MHLPSPTDSKFYRTLMEBEDMEDI VDAEYLVPHQGF-----NSPST- 454
Qy 1048 SGGGDLTLGLEPSEEAAPRSP-----APSEGAGSDVDFDGLMGAAKGLQLSPLTHDPS 1102
Db 455 -----SRTPLLSLSATSNSATNCID-----RNGQGHVREDSP 489
Qy 1103 LQRYSEDPTVPLPSET--DGYVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPGAT 1160
Db 490 VQRYSSDPTGNFLEESIDDGFL-----PAPEYVNO--LMPKKPS----- 526
Qy 1161 LERAKTLPFGKGVKVDVF-----AFGGAVENPEYL 1191
Db 527 -----TAMVQNIYNNISLTAISKLPMDSRVQNSHSTAVDNPEYL 566

RESULT 13
Q64895
ID Q64895 PRELIMINARY; PRT; 962 AA.
AC Q64895;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Gag, v-erb-A, v-erb-B protein.
GN Gag, v-ERB-A, v-ERB-B.
OS Avian erythroblastosis virus.
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90206603; PubMed=1969616;
RA Bruskina A., Jackson J., Bishop J.M., McCarley D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the
RT transforming potential of the oncogene v-erb-B.";
```



```
RL Oncogene 5:15-24 (1990).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; X52209; CAA36459.1; -.
DR EMBL; X52211; CAA36459.1; JOINED.
DR HSSP; P10828; 2NLL.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Steroid_receptor.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD000001; Euk_pkinase; 1.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00219; TyrKc; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW Transcription regulation; Transferase; Tyrosine-protein kinase;
KW Zinc-finger.
SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791E4E95CE CRC64;

Query Match 22.8%; Score 1543.5; DB 15; Length 962;
Best Local Similarity 49.6%; Pred. No. 3.3e-108;
Matches 343; Conservative 72; Mismatches 152; Indels 125; Gaps 19;

Qy 541 VEECRVLQGLPRE-YVNAR-HCLP-----CHPEQ 568
Db 354 IEKQESYLLAFEHYINRYKHNIHFWSKLMKVADLRMTGAYHASRFLMKVECPTELS 413
Qy 569 PQNGSVTCFGEADQCVACAHYKDPFCVACRCPGVKPDLSYMPIMKFFPDEEGACQPCPI 628
Db 414 PQE-----VGP--DHCWKCAHFIDGPHCVKACPVGLGENDTL-VKYYADANAVCOLCHP 465
Qy 629 NCHTSCVDLDDKGPASORASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQKIRKYTM 687
Db 466 NCTRGCKGPGLEGCP--NGSKTPSIAAGVVGGLLCVVGLGIGLYLRRR-HIVKRTL 521
Qy 688 RLLQETELVEPLTPSGAMPNOAQRILKETELRKVKVGLSGAFQYIK-----ANSKF 741
Db 522 RLLQERLVEPLTPSGEAPNQAHRLILKETEFKKVKVGLGFGAFGTYYKGLWIPEGSKVT 581
Qy 742 IGITELVLRNTSPKANKEILDEAYVMAGVSPVSRLLGICLTSTVQLVTQMLPYGCLL 801
Db 582 IPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQMLPYGCLL 641
Qy 802 DHVENRGRGSDQLLNCMQIAKMSYLEDLVRLHRLDAARNLVKSPNHRKITDFGLA 861
Db 642 DYIREHKDNGISQYLLNWCQIAKGMNLYEERHVRDLAARNLVKTPQHVKITDFGLA 701
Qy 862 RLLDIDETEYHAGCKVPIKMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIP 921
Db 702 KQLGADEKEYHAEGGKVPKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIP 761
Qy 922 ARETPDLLEKGERLPQPPICITIDVYMWKMWIDSECRPRFRELVSERFARMARDPQFV 981
Db 762 ASELSVLEKGERLPQPPICITIDVYMWKMWGADSRPKFRELIAEFKSMARDPPRYL 821
Qy 982 VIQ-NEDLGPASPLDSTFYRSLLEDGDLVDAEYLVFQQGFPFCDPAPFAGMGVHHR 1040
Db 822 VIQDERMHLPSPTDSKFYRLTMEEDMEDIVDAEYLVFHQGF-----866
Qy 1041 HRSSTSGGDDLTLGLEPSEEEAPRSPPLAPSEGAGSDVFDGLGMGAAGKLSLPHDHP 1100
Db 867 -NSPST-----SRTPLLSSLSATSN-----NSATKCIDRNGGH-- 898
```

```
Qy 1101 SPLQRYSEDPVLPSETDGYVAPLTCSPQPEYVNPQDVRPQPSPREGPLPAARFAGAT 1160
Db 899 -----PVREDGFL-----PAPEYVNO--LMPKKPSTAMVQNIYVVISLT 936
Qy 1161 -LERAKTLSPGKNGVVDVFAFGGAVENPEYL 1191
Db 937 AISKLPMDSFYQN-----SHSTAVDNPEYL 961

RESULT 14
Q85468 PRELIMINARY; PRT; 545 AA.
ID Q85468
AC Q85468;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Avian Erythroblastosis virus (IEB34) v-erbB gene.
OS Avian erythroblastosis virus.
OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88217326; PubMed=2897102;
RA Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
RT virus mutants that are temperature sensitive for transformation.";
RL Oncogene Res. 1:265-278(1987).
DR ENBL; X06943; CAA30024.1; -.
DR HSSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 545 AA; 60899 MW; 140DCE8CCA0F8AF4 CRC64;

Query Match 22.7%; Score 1535; DB 15; Length 545;
Best Local Similarity 52.5%; Pred. No. 6.2e-108;
Matches 330; Conservative 69; Mismatches 131; Indels 98; Gaps 16;

Qy 578 GPEADOCVACAHYKDPFCVACRCPGVKPDLSYMPIMKFFPDEEGACQPCINCHTSCVDL 637
Db 1 GP--DHCWKCAHFIDGPHCVKACPVGLGENDTL-VKYYADANAVCOLCHPNCRTGCKGP 57
Qy 638 DKGCPAEQASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQKIRKYTMRELLQETEL 696
Db 58 GLEGCP--NGSKTPSIAAGVVGGLLCVVGLGIGLYLRRR-HIVKRTLRELLQEREL 113
Qy 697 VEPLTPSGAMPNOAQRILKETELRKVKVGLSGAFQYIK-----ANSKFIGITELVLR 750
Db 114 VEPLTPSGEAFNQAHRLILKETEFKKVKVGLGFGAFGTYYKGLWIPEGKVTIPVAIKELR 173
Qy 751 ENTPSPKANKEILDEAYVMAGVSPVSRLLGICLTSTVQLVTQMLPYGCLLDHVENRGR 810
Db 174 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQMLPYGCLLDYIREHKON 233
Qy 811 LGSQDILLNCMQIAKMSYLEDLVRLHRLDAARNLVKSPNHRKITDFGLARLLDIDETE 870
Db 234 IGSQYLLNWCQIAKGMNLYEERHVRDLAARNLVKTPQHVKITDFGLAKQLGADEKE 293
Qy 871 YHADGCKVPIKMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLE 930
Db 294 YHAEGCKVPIKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISSVLE 353
Qy 931 KGERLPQPPICITIDVYMWKMWIDSECRPRFRELVSERFARMARDPQFVVIQ-NEDLIG 989
Db 354 KGERLPQPPICITIDVYMWKMWIDSECRPRFRELIAEFKSMARDPPRYLVYIQGDERMH 413
```

```

Qy 990 PASPLDSTFYRSLLEDDMDGLVDAAEYLVPOQGFCDPAPGAGMVHRRSSSTRSG 1049
Dl 414 LPSFTDSKFYTLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
Qy 1050 GGDLTGLGLEPSEEARPSPL-----APSEGAGSDVFDGDLGMAAKGLQSLPTHDRPSLQ 1104
Dl 455 -----SRTPLLSLSATSNSNATNCIDRNG-----H----- 481
Qy 1105 RYSEDPTVLPSETDGVVAPLTCSPQEPYVQNPQVRPQPSREGPLPAARPAGAT-LE 1163
Dl 482 -----PVRDGL-----PAEYVQ-----LMPKPKSTAMVQIQVYISLTAISK 523
Qy 1164 AKTLSPGKGVYKDVFAFGGAVENPEYL 1191
Dl 524 LPMDSRYQN-----SHSTAVDNPPEYL 544

RESULT 15
Q9WVF5 PRELIMINARY; PRT; 655 AA.
AC Q9WVF5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor
DE isoform 3).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Thredgill D.W., Danielson A.J., Schell C.,
RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Maible N.J.;
RT "Alternative Transcripts from the Human and Mouse EGFR Genes Encode
RT Carboxy-Terminal Truncated Receptors.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTC;
RA Kawai J., Thredgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sinclair C.S., Pearse R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Iyoy-oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohteki S.,
RA Hayashizaki Y.;

```

```

RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690 (2001).
DR EMBL; AF124513; AAD44149.1; -
DR EMBL; AF275366; AAG28047.1; -
DR EMBL; AF275364; AAG28047.1; JOINED.
DR EMBL; AF275365; AAG28047.1; JOINED.
DR EMBL; AK004944; BAB23688.1; -
DR EMBL; AK004883; BAB23641.1; -
DR EMBL; AK004911; BAB23662.1; -
DR MGI; MGI:95294; Egfr.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; FU_3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match 22.2%; Score 1506.5; DB 11; Length 655;
Best Local Similarity 44.5%; Pred. No. 1.2e-105;
Matches 285; Conservative 97; Mismatches 233; Indels 25; Gaps 9;

Qy 11 LLALLPPGAA--STOVCTGDMKRLPASPETHLDMLRHLHYGGCVQVGNLELTLYPTN 68
Dl 14 LLTALCAAGGAALEKKVCOQTSNRLTQLGTFEDHFLSLQRMVNNCEVLGNLEITVQRN 73
Qy 69 ASLSFLQDIOEVQGVYLIHNOVQVPLQRLRVGCTOLFEDNYALVLDNGDPLNNTTP 128
Dl 74 YDLSFLKTIOEVAGYVLIALTVERIPLENLQIRNALYENTYALAILSN----- 134
Qy 129 VTGASPGGLRELQRLSLTEILKGGVLIQNPOLCYQDTILWKDI-----FKNNQLALTLI 184
Dl 125 -YGTNETGLRELPMRNQLIILGAVRFSNNPILCNMDTIQWRDIVQNVFMSNMSMDL--- 180
Qy 185 DYNRSRACHPCSPMKSGKSCWGESSEDCQSLRTTVCAAGCA-CKGPLPTDCHEQCAAG 243
Dl 181 -QSHPSCKPCPCSPGNGSGGEECQKLTIIQAQCSHRGRSPSDCCCHNQCAAG 239
Qy 244 CTGPKHSDCLCLHFNHSGICELHCPALVYNTDTFESMPNPEGRVTCGASCVTACPNY 303
Dl 240 CTGPRESDCLVCKQFQDEATCKDTCPLMLYNTTYQMDVNPPEGRYSFGATCVKKCPNY 299
Qy 304 LSTDVSGSCTLVCPHMQEVTAEQTCRCKSKPCARVCVGLGMQYIKANSKEFIGITELE 363
Dl 300 VYTDHSGSVRACGPDYEV-EEDGIRKCKKCGPCRKVCNGIGIGEFK-DTLSINATNIK 357
Qy 364 -FAGCKKIFGSLAFIPESFDGDPASNTAPLQEPQLQVFTLEITCYLYISAWPDSLPL 422
Dl 358 HFKYCTAISGDLHILPFAFKGDSFTTPPLDPRELEILKTVKEITGFLLIQAMPDNTDL 417
Qy 423 SVFQNLQVIRGLTHNGAYSLTLQGLISWLGRLSELGSLGLALHNNHLCFVHTVPW 482
Dl 418 HAFENLEIRGRTKQHGQFSLAVVGNLITSLGRLSLKESDGDVIIISGRNRLCYANTINW 477
Qy 483 DQLFRPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPPTQCVNCSQFLRGQECVE 542
Dl 478 KKLFGTPTNKTMMNRAEKCKAVNHNVCNPLCSSEGCWGPPEPRDCVSCQVNSRGECVE 537
Qy 543 ECKVLGGLPREYVNAHCLPCHPEQOPQNGSVTCFGEADQCACAHYKDPPEFCVACRPS 602
Dl 538 KCNILEGEPREFVENSECICQHPCECLPQAMNITCTGRPDNCICQAHYIDGPHCVKTCPA 597
Qy 603 GVKPDLISYMPIWKFPEEGACQPCPINCETHSCVDLDDKGC 642
Dl 598 GINGENNTLI-VWKYADANNVCHLCHANCTYGCAGPGLQGC 636

```

Search completed: July 22, 2003, 09:24:45
Job time : 51.6008 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:41:54 ; Search time 36.7573 Seconds
(without alignments)
4527.811 Million cell updates/sec

Title: SEQ4-710-730-12

Perfect score: 6775

Sequence: 1 MELAALCRWGLLALLPPGA.....TFKGTPTAENPEYLGLDVVPV 1249

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6608	97.5	1255	21	Human heregulin 2
2	6608	97.5	1255	22	Human tyrosine kin
3	6608	97.5	1255	22	HER2 transgene pla
4	6608	97.5	1255	23	Human HER2 (ErbB2)
5	6602	97.4	1255	17	HER-2/neu protein.
6	6602	97.4	1255	20	Human HER-2/neu on
7	6602	97.4	1255	21	Human HER-2/neu pr
8	6602	97.4	1255	21	Amino acid sequenc
9	6602	97.4	1255	22	Human HER-2/neu pr
10	6602	97.4	1255	22	HER2/neu amino aci

11	6602	97.4	1255	23	AAE24067	Human Her-2 protei
12	6602	97.4	1255	23	AAE20479	Human Her-2/neu pr
13	6602	97.4	1255	23	AAE51143	Human Her-2/neu on
14	6602	97.4	1255	23	AAU77114	Human Her-2/neu po
15	6559	96.8	1433	14	AAE39568	Sequence of c-erbB
16	6438	95.0	1223	23	AAU98923	Human breast cance
17	6285	92.8	1200	21	AAE21208	Human HER-2/neu pr
18	5815.5	85.8	1256	21	AAE21199	Rat HER-2/neu prot
19	5815.5	85.8	1256	23	AAE51144	Rat HER-2/neu onco
20	5788.5	85.4	1256	21	AAE21206	Mouse Her-2/neu pr
21	5788.5	85.4	1256	22	AAE62860	Amino acid sequenc
22	5788.5	85.4	1256	23	AAE51151	Mouse Her-2/neu on
23	4817	71.1	919	21	AAE21203	Human HER-2/neu fu
24	4817	71.1	919	23	AAE51148	Her-2/neu extracel
25	4067.5	60.0	920	23	AAE51152	Mouse Her-2/neu ex
26	4067.5	60.0	926	23	AAE51153	Mouse Her-2/neu ex
27	3701	54.6	712	21	AAE21204	Human HER-2/neu fu
28	3701	54.6	712	23	AAE51149	Her-2/neu extracel
29	3552	52.4	782	18	AAW19764	Her2-GM-CSF immuno
30	3550	52.4	653	21	AAE21200	Extracellular HER-
31	3550	52.4	653	23	AAE51145	Human Her-2/neu on
32	3512	51.8	645	22	AAE60408	Human ErbB2 oncopr
33	3512	51.8	645	22	AAE61593	Human ErbB2 extrac
34	3447	50.9	951	21	AAE44393	DC8scFv-erbB2EC fu
35	3344	49.4	624	11	AAE08222	Extracellular port
36	3053.5	45.1	654	21	AAE21205	Rat HER-2/neu prot
37	3053.5	45.1	654	23	AAE51150	Rat HER-2/neu onco
38	3030	44.7	1210	21	AAE19259	Amino acid sequenc
39	3030	44.7	1210	21	AAE50616	Human EGF receptor
40	3030	44.7	1210	23	AAE23019	Human Her-1 protei
41	3028	44.7	1210	23	AAE50768	Human epidermal gr
42	3028	44.7	1210	22	AAE68420	Amino acid sequenc
43	2989	44.1	1210	23	ABP51768	Human epidermal gr
44	2958	43.7	583	23	AAE20483	Human protein for
45	2958	43.7	587	23	AAE20481	Human protein for

ALIGNMENTS

RESULT 1
AAE24067
ID AAE24067 standard; Protein; 1255 AA.
XX AAE24067;
XX AAE24067;
DT 10-AUG-2000 (first entry)
XX Human heregulin 2 (Her2).
XX
KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Domain 1..173
XX /label= N-terminal
XX /note= "mature polypeptide"
XX Region 5..25
XX /label= insertion region
XX /note= "suitable for foreign epitope insertion"
XX Region 59..73
XX /label= insertion region
XX /note= "suitable for foreign epitope insertion"
XX Region 103..117
XX /label= insertion region
XX /note= "suitable for foreign epitope insertion"
XX Region 149..163
XX /label= insertion region
XX /note= "suitable for foreign epitope insertion"
XX Domain 174..323

FT Region /label= Cysteine_rich_domain
 FT 210..224
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 250..264
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 324..483
 FT /label= Ligand_binding_domain
 FT 325..339
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 369..383
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 465..479
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 484..623
 FT /label= Cysteine_rich_domain
 FT 579..593
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 624..654
 FT /label= Transmembrane_domain
 FT 632..652
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 653..667
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 655..1010
 FT /label= Tyrosine_kinase_domain
 FT 661..675
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 695..709
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 710..730
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 1011..1235
 FT /label= C-terminal_domain
 XX WO200020027-A2.
 XX 13-APR-2000.
 XX 05-OCT-1999; 99WO-DK00525.
 XX 05-OCT-1998; 98DK-0001261.
 XX 20-OCT-1998; 98US-0105011.
 XX (MEEI-) M & E BIOTECH AS.
 XX Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
 PI Gautam A, Birk P, Karlsson G;
 PI WPI; 2000-349917/30.
 DR N-PSDB; AAA09455.
 XX Inducing immune responses to weakly immunogenic, tumor associated
 XX peptide antigens for the treatment of breast and prostate cancer
 XX Claim 62; Page 193-198; 220pp; English.
 XX This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
 XX Her2 can be used in the claimed method as an autovaccine to induce a CTL
 XX response. Subdominant CTL epitopes, antibody binding regions and
 XX cysteine residues involved in disulfide bonds are preserved in the
 XX immunogenized forms. Regions suitable for the insertion of foreign T
 XX helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
 CC cell-associated peptide antigens (PA) such as those associated with
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
 CC The method comprises effecting simultaneous presentation by antigen
 CC producing cells (APCs) of the animals immune system of: (1) at least 1
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
 CC B-cell group derived from the cell-associated PA; and (2) at least 1
 CC first T-helper cell group which is foreign to the animal. Analogues of
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
 CC part of all known and predicted CTL and B-cell epitopes of the respective
 CC PA and including at least one foreign T helper epitope are also claimed.
 CC The method is used to treat prostate, prostate/breast or breast cancer
 CC when the PA is human PSM, FGF8b and Her2, respectively.
 XX
 SQ Sequence 1255 AA;
 Query Match 97.5%; Score 6608; DB 21; Length 1255;
 Best Local Similarity 97.4%; Pred. No. 0;
 Matches 1223; Conservative 5; Mismatches 20; Indels 8; Gaps 2;
 Qy 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLPASPETHLDMLRLHLYQGCVVQGNL 60
 Db |||||
 Qy 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLPASPETHLDMLRLHLYQGCVVQGNL 60
 Db |||||
 Qy 61 ELTYLPTNASLSFLQDIOEVQGVVLIAHNOVROVPLQRLRIVRGTOLEFEDNYALAVLDNG 120
 Db |||||
 Qy 61 ELTYLPTNASLSFLQDIOEVQGVVLIAHNOVROVPLQRLRIVRGTOLEFEDNYALAVLDNG 120
 Db |||||
 Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILMKDIFHNKQLA 180
 Db |||||
 Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILMKDIFHNKQLA 180
 Db |||||
 Qy 181 LTLIDNRSRACHPCSPMKGSRGWGSESSDCOSLTRTVCCAGCACRCKGLPTDCCHEQC 240
 Db |||||
 Qy 181 LTLIDNRSRACHPCSPMKGSRGWGSESSDCOSLTRTVCCAGCACRCKGLPTDCCHEQC 240
 Db |||||
 Qy 241 AACCTGPKSHDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Db |||||
 Qy 241 AACCTGPKSHDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Db |||||
 Qy 301 YNYLSTDVGSCTLVCPPLHNOVETAEQGTQRCCKSPCARVCYGLGMYIKANSKFIT 360
 Db |||||
 Qy 301 YNYLSTDVGSCTLVCPPLHNOVETAEQGTQRCCKSPCARVCYGLGMYIKANSKFIT 360
 Db |||||
 Qy 361 ELEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLOVPETLEEITGYLIYISAWPDSL 420
 Db |||||
 Qy 361 IQEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLOVPETLEEITGYLIYISAWPDSL 420
 Db |||||
 Qy 421 DLSVFONLQVIRGRILHNGAYSILTLQGLISWGLSLRELGSGLALIHNTLHLCFVHTV 480
 Db |||||
 Qy 421 DLSVFONLQVIRGRILHNGAYSILTLQGLISWGLSLRELGSGLALIHNTLHLCFVHTV 480
 Db |||||
 Qy 481 PWDQLFRNPQALLHTANRPEDECVGEGGLACHOLCARGHGWGPGTQCVCNCSQFLRGQEC 540
 Db |||||
 Qy 481 PWDQLFRNPQALLHTANRPEDECVGEGGLACHOLCARGHGWGPGTQCVCNCSQFLRGQEC 540
 Db |||||
 Qy 541 VEESCRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFPCVARC 600
 Db |||||
 Qy 541 VEESCRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFPCVARC 600
 Db |||||
 Qy 601 PSQVGPDLSPYMPIWKPDPDEGACQPCINCTHSCVDLDDKGCPCAEORASPLTISVSAVVG 660
 Db |||||
 Qy 601 PSQVGPDLSPYMPIWKPDPDEGACQPCINCTHSCVDLDDKGCPCAEORASPLTISVSAVVG 660
 Db |||||
 Qy 661 ILLVVLGVVFGILIKRRQOKIRKYTMRLLOTELVEPLTPSGAMPNQAMRILKETEL 720
 Db |||||
 Qy 661 ILLVVLGVVFGILIKRRQOKIRKYTMRLLOTELVEPLTPSGAMPNQAMRILKETEL 720
 Db |||||
 Qy 721 RKVKVLGSGAFQGYIYA-----NSKFTGITELVLRNTSPKANKEILDEATVMAGVGS 773
 Db |||||
 Qy 721 RKVKVLGSGAFQGYIYA-----NSKFTGITELVLRNTSPKANKEILDEATVMAGVGS 773
 Db |||||

Qy 774 PYVSRLLGICLTSTVQLVQMPYGCCLLDHVRNRRGLSGDILLNWCQIAKGMVLEDDV 833
Db 780 PYVSRLLGICLTSTVQLVQMPYGCCLLDHVRNRRGLSGDILLNWCQIAKGMVLEDDV 839
Qy 834 RLVRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNWALESIILRRRF 893
Db 840 RLVRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNWALESIILRRRF 899
Qy 894 THQSDVMSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYIMVKCW 953
Db 900 THQSDVMSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYIMVKCW 959
Qy 954 MIDSECPRRFELVSEFSRMARDQRFVVIQNEIDLGASPLDSTFYRSLLDDDDMDGLVD 1013
Db 960 MIDSECPRRFELVSEFSRMARDQRFVVIQNEIDLGASPLDSTFYRSLLDDDDMDGLVD 1019
Qy 1014 ABEYLVPQQGFCCPDAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEEAPRSPAPSE 1073
Db 1020 ABEYLVPQQGFCCPDAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEEAPRSPAPSE 1079
Qy 1074 GAGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEY 1133
Db 1080 GAGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEY 1139
Qy 1134 VNQPDVRPQPPSPREGPLPAARPAAGATLERAKTUSPGKNGVVKDVFAFGGAVENPEYLT 1193
Db 1140 VNQPDVRPQPPSPREGPLPAARPAAGATLERAKTUSPGKNGVVKDVFAFGGAVENPEYLT 1199
Qy 1194 QGGAAPQHPHPPAFSPAFNDLYWDQPPERGAAPPSTFKGTPTAENPEYLGLDVVP 1249
Db 1200 QGGAAPQHPHPPAFSPAFNDLYWDQPPERGAAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 2
AAE12130
ID AAE12130 standard; Protein; 1255 AA.
XX
AC AAE12130;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human tyrosine kinase-type receptor, HER-2.
XX
KW Therapeutic compound; major histocompatibility complex; vaccine;
KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
KW antigen presenting cell; human; tyrosine kinase-type receptor.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 774..782
FT /note="Antigenic epitope"
XX
PN WO200168677-A2.
XX
PD 20-SEP-2001.
XX
PF 16-MAR-2001; 2001WO-US40328.
XX
PR 16-MAR-2000; 2000US-0527487.
XX
PA (GENZ) GENZYME CORP.
XX
PI Nicolette CA;
XX
DR WPI; 2001-616284/71.
DR N-PSDB; AAD19731.
XX
PT Novel synthetic therapeutic compound for inducing immune response and
PT for use in adoptive immunotherapy, has enhanced binding to major
PT histocompatibility molecules and enhanced immunoregulatory properties

XX

Claim 4; Page 63-67; 69pp; English.

XX
PS The invention relates to synthetic therapeutic compounds (antigenic
CC peptides) with enhanced binding to major histocompatibility complex
CC (MHC) molecules and enhanced immunoregulatory properties relative
CC to their natural counterparts. Compounds of the invention are useful
CC for inducing an immune response in a subject and for use in adoptive
CC immunotherapy. They are useful as components of anti-cancer vaccines
CC and to expand immune effector cells that are specific for cancers
CC characterised by expression of the breast cancer antigen, HER-2.
CC Polynucleotides that encode peptides of the invention are useful as
CC hybridisation probes and as primers for the detection of genes of gene
CC transcripts that are expressed in antigen presenting cells (APCs), to
CC confirm transduction of polynucleotides into host cells. The present
CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
CC of the invention are designed based on the HER-2 antigenic peptide
CC (774-782).

SQ Sequence 1255 AA;

Query Match 97.5%; Score 6608; DB 22; Length 1255;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 1223; Conservative 5; Mismatches 20; Indels 8; Gaps 2;

Qy 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKRLPASPETHLDMRLHLYQGCVVQGNL 60

Db 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKRLPASPETHLDMRLHLYQGCVVQGNL 60

Qy 61 ELTYLPTNASLSFLQDIOEVQGVYLIAHNQVROVPLQRLRIVRGTLFEDNYALAVLDNG 120

Db 61 ELTYLPTNASLSFLQDIOEVQGVYLIAHNQVROVPLQRLRIVRGTLFEDNYALAVLDNG 120

Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTTLWKDI FHKNNOLA 180

Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTTLWKDI FHKNNOLA 180

Qy 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCCAGCARCKGLPTDCCHEQC 240

Db 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCCAGCARCKGLPTDCCHEQC 240

Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy 301 YNYLSTDVSGSTLVCPLHNOEVTAEADGTQCEKSCPCARVCYGLGMOYIKANSKFIT 360

Db 301 YNYLSTDVSGSTLVCPLHNOEVTAEADGTQCEKSCPCARVCYGLGMOYIKANSKFIT 360

Qy 361 ELEFAGCKKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFPETLEEITGYLYISAWPDSL 420

Db 361 IQEFAGCKKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFPETLEEITGYLYISAWPDSL 420

Qy 421 DLSVFONLQVIRIILHNGAYSLTLOGLISWLGSLRELGSGLALIHNTLHCFVHTV 480

Db 421 DLSVFONLQVIRIILHNGAYSLTLOGLISWLGSLRELGSGLALIHNTLHCFVHTV 480

Qy 481 PWDOLFNRPHCALHTANRPEDECVGEGGLACHOLCARGHCWGPGTQCVNCSQFLRGQEC 540

Db 481 PWDOLFNRPHCALHTANRPEDECVGEGGLACHOLCARGHCWGPGTQCVNCSQFLRGQEC 540

Qy 541 VEESCRVLQGLPREYVNNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFVCVARC 600

Db 541 VEESCRVLQGLPREYVNNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFVCVARC 600

Qy 601 PSGVKPDLSPYMPIWKFPDEBEGACQPCPINCTHSCVDLDDKGCPAERASPLTSIVSAVVG 660

Db 601 PSGVKPDLSPYMPIWKFPDEBEGACQPCPINCTHSCVDLDDKGCPAERASPLTSIVSAVVG 660

Qy 661 ILLVAVLVGVVFGILIKRROOKIRKYTMRLLOETELVEPLTPSGAMPNQAQWILKETEL 720

Db 661 ILLVAVLVGVVFGILIKRROOKIRKYTMRLLOETELVEPLTPSGAMPNQAQWILKETEL 720

721 RKVKVLSGAGFGQYIKA-----NSKFIGITELVRENTSPKANKEILD EAYVMAGVGS 773
 721 RKVKVLSGAGFGTVYKGIWIPDGENVK-IPVAIKVLRNTSPKANKEILD EAYVMAGVGS 779
 774 PYVSRLLGICLTSTVQLVTQMPYGCCLLDHVRENRGRIGSDLLNWCQIAKMSYLEDV 833
 780 PYVSRLLGICLTSTVQLVTQMPYGCCLLDHVRENRGRIGSDLLNWCQIAKMSYLEDV 839
 834 RLVRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALESI LRRRF 893
 840 RLVRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALESI LRRRF 899
 894 THQSDVMSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPPOPICTIDVYIMVKW 953
 900 THQSDVMSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPPOPICTIDVYIMVKW 959
 954 MIDSECRPRFRELVSERPMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDMGDLVD 1013
 960 MIDSECRPRFRELVSERPMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDMGDLVD 1019
 1014 AEEYLVPOQGFCDPAPAGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSE 1073
 1020 AEEYLVPOQGFCDPAPAGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSE 1079
 1074 GAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEY 1133
 1080 GAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEY 1139
 1134 VNQSDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAPGGAVENPEY LTP 1193
 1140 VNQSDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAPGGAVENPEY LTP 1199
 1194 OGGAAPQHPHPPAPSPAFDNLVYWDQDPPERGAAPPSTFKGPTAENPEYLGDLV 1249
 1200 OGGAAPQHPHPPAPSPAFDNLVYWDQDPPERGAAPPSTFKGPTAENPEYLGDLV 1255
 RESULT 3
 AAB60167
 ID AAB60167 standard; Protein; 1255 AA.
 AC AAB60167;
 XX
 XX
 DT 03-APR-2001 (first entry)
 XX
 DE HER2 transgene plasmid construct encoded protein.
 XX
 KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
 KW antibody.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 FN WO200100244-A2.
 XX
 XX 04-JAN-2001.
 PD
 XX
 PF 23-JUN-2000; 2000WO-US17229.
 XX
 PR 25-JUN-1999; 99US-0141316.
 PR 16-MAR-2000; 2000US-0189844.
 XX
 XX (GETH) GENENTECH INC.
 PA
 XX
 FI Erickson S, Schwall R;
 XX
 DR WPI; 2001-061962/07.
 DR N-PSDB; AAF24297.
 XX
 XX
 PT Treating tumors, particularly breast cancers, which overexpress an ErbB
 PT receptor and does not respond to an anti-ErbB antibody, comprises
 PT conjugating the antibody to a maytansinoid -
 XX

Example 3; Fig 4; 92pp; English.

PS The present invention provides a method of treating cancer by
 XX administering a conjugate of anti-ErbB antibody with a maytansinoid. In
 CC particular, the antibody is directed against ErbB2 (also known as HER2
 CC and p185neu). The method is particularly useful in the treatment of
 CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
 CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
 XX
 SQ Sequence 1255 AA;

Query Match 97.5%; Score 6608; DB 22; Length 1255;
 Best Local Similarity 97.4%; Pred. No. 0;
 Matches 1223; Conservative 5; Mismatches 20; Indels 8; Gaps 2;
 QY 1 MELAALCRWGLLALLPAGAASTQVCTGDKMLRLPASPTHLDMLRHL YQGCVVQGNL 60
 DB 1 MELAALCRWGLLALLPAGAASTQVCTGDKMLRLPASPTHLDMLRHL YQGCVVQGNL 60
 QY 61 ELYLPTNASLSFLQDIQEVQGVVLIHNVORVPLQRLRIVRGTO L FEDNYALAVLDNG 120
 DB 61 ELYLPTNASLSFLQDIQEVQGVVLIHNVORVPLQRLRIVRGTO L FEDNYALAVLDNG 120
 QY 121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRPOLCYQDTILWKDI FHNQOLA 180
 DB 121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRPOLCYQDTILWKDI FHNQOLA 180
 QY 181 LTLIDNRSRACHPCSPMKSGRCWGESSEDCOSLRTVTCAGGCARCKG LPTDCCHEQC 240
 DB 181 LTLIDNRSRACHPCSPMKSGRCWGESSEDCOSLRTVTCAGGCARCKG LPTDCCHEQC 240
 QY 241 AAGCTGPKSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 DB 241 AAGCTGPKSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 QY 301 YNYLSTDVSGCTIVCP LHNQEVTAEDGTQRCCKSPCARVCYGLGQYIKANSKFITG 360
 DB 301 YNYLSTDVSGCTIVCP LHNQEVTAEDGTQRCCKSPCARVCYGLGQYIKANSKFITG 360
 QY 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVETLEEITGYLIYI SAWPDSLP 420
 DB 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVETLEEITGYLIYI SAWPDSLP 420
 QY 421 DLSVFQNLQVIRGRIHNGAYSLTLOGLGISWLGSLRSLRSLGSLALIHNT HLCFVHTV 480
 DB 421 DLSVFQNLQVIRGRIHNGAYSLTLOGLGISWLGSLRSLRSLGSLALIHNT HLCFVHTV 480
 QY 481 PWDLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
 DB 481 PWDLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
 QY 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFVCVARC 600
 DB 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFVCVARC 600
 QY 601 PSGVKPDLSPYMPIWKFPDEBACQPCPINCTHSCVDLDDKGC PAEQRASPLTSIVSAVVG 660
 DB 601 PSGVKPDLSPYMPIWKFPDEBACQPCPINCTHSCVDLDDKGC PAEQRASPLTSIVSAVVG 660
 QY 661 ILLVVLVGVVFGILIKRQKIRKYTMRRLLQTELVEPLTPSGAMPNQAMRILKETEL 720
 DB 661 ILLVVLVGVVFGILIKRQKIRKYTMRRLLQTELVEPLTPSGAMPNQAMRILKETEL 720
 QY 721 RKVKVLSGAGFGQYIKA-----NSKFIGITELVRENTSPKANKEILD EAYVMAGVGS 773
 DB 721 RKVKVLSGAGFGTVYKGIWIPDGENVK-IPVAIKVLRNTSPKANKEILD EAYVMAGVGS 779
 QY 774 PYVSRLLGICLTSTVQLVTQMPYGCCLLDHVRENRGRIGSDLLNWCQIAKMSYLEDV 833
 DB 780 PYVSRLLGICLTSTVQLVTQMPYGCCLLDHVRENRGRIGSDLLNWCQIAKMSYLEDV 839
 QY 834 RLVRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALESI LRRRF 893
 DB 834 RLVRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALESI LRRRF 893

Db 840 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILRRRF 899
 QY 894 THQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYIMVKCW 953
 Db 900 THQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYIMVKCW 959
 QY 954 MIDSECPRELVSESRMARDQRFVVIQNEDLGPASPLDSTFYRSLLDDDDMGDLVD 1013
 Db 960 MIDSECPRELVSESRMARDQRFVVIQNEDLGPASPLDSTFYRSLLDDDDMGDLVD 1019
 QY 1014 ABEYLVPQGGFFCDPAPGAGGMVHRRSSRSGGDLTLGLEPSEERAPRSLAPSE 1073
 Db 1020 ABEYLVPQGGFFCDPAPGAGGMVHRRSSRSGGDLTLGLEPSEERAPRSLAPSE 1079
 QY 1074 GAGSDVFDGLGMAAGKGLQSLPHTDPSPLQRYSEDTVPPLSETDGYVAPLTCSPQPEY 1133
 Db 1080 GAGSDVFDGLGMAAGKGLQSLPHTDPSPLQRYSEDTVPPLSETDGYVAPLTCSPQPEY 1139
 QY 1134 VNQPDVTPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVDVFAFGGAVENPEYLT 1193
 Db 1140 VNQPDVTPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVDVFAFGGAVENPEYLT 1199
 QY 1194 QGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGDLV 1249
 Db 1200 QGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGDLV 1255

RESULT 4
 AAU74545
 ID AAU74545 standard; Protein; 1255 AA.
 XX AAU74545;
 DT
 XX
 XX 23-APR-2002 (first entry)
 DE Human HER2 (ErbB2) polypeptide.
 XX
 KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;
 KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
 KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
 KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
 KW glial disorder; astrocytal disorder; hypothalamic disorder;
 KW glandular disorder; macrophagal disorder; epithelial disorder;
 KW stromal disorder; blastocoealic disorder; inflammatory disorder;
 KW angiogenic disorder; immunological disorder.
 XX
 OS Homo sapiens.
 XX
 PN US2002001587-A1.
 XX
 PD 03-JAN-2002.
 XX
 PF 16-MAR-2001; 2001US-0811123.
 XX
 PR 16-MAR-2000; 2000US-189844P.
 PR 05-OCT-2000; 2000US-238327P.
 XX
 XX (ERIC/) ERICKSON S.
 PA (SCHW/) SCHWALL R.
 PA (SLIW/) SLIWOWSKI M.
 XX
 XX Erickson S, Schwall R, Sliwowski M;
 XX WPI; 2002-163686/21.
 XX
 DR N-PSDB; ABK14058.
 XX
 XX Treating tumour characterised by overexpression of epidermal growth
 PT factor receptor, ErbB or cancer in mammal, comprises administering
 PT anti-ErbB antibody-maytansinoid conjugate to the mammal -
 XX
 XX Example 3; Fig 7; 93pp; English.
 PS
 XX The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor
 CC receptor (ErbB) and does not respond or responds poorly, to treatment
 CC with an anti-ErbB antibody, comprising administering to the mammal an
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
 CC prostate and bladder, preferably breast cancer. The breast cancer is a
 CC metastatic breast cancer or an aggressive form of metastatic breast
 CC cancer which overexpresses ErbB2. The method is also useful for treating
 CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
 CC epithelial stromal, blastocoealic, inflammatory, angiogenic and
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)
 CC polypeptide of the invention.

XX
 SQ Sequence 1255 AA;

Query Match 97.5%; Score 6608; DB 23; Length 1255;

Best Local Similarity 97.4%; Pred. No. 0;
 Matches 1223; Conservative 5; Mismatches 20; Indels 8; Gaps 2;

QY 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKRLRLPASPEHLDMRLHLYQGQVVOG 60
 Db 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKRLRLPASPEHLDMRLHLYQGQVVOG 60
 QY 61 ELYLPTNASLSPLQDIOEVQGVLIHQNQVROVPLQRLIRVGTOLFDNYALAVLDNG 120
 Db 61 ELYLPTNASLSPLQDIOEVQGVLIHQNQVROVPLQRLIRVGTOLFDNYALAVLDNG 120
 QY 121 DPLNNTTPTVGTASPGGLRELRLTEILKGGVLIQORNPOLCYODTILWKDIFHNKOLA 180
 Db 121 DPLNNTTPTVGTASPGGLRELRLTEILKGGVLIQORNPOLCYODTILWKDIFHNKOLA 180
 QY 181 LTLIDNTRSRACHPCSPMKGSRWGESSEDCOSLTRTVCAGGCARCKGLPTDCCHEQC 240
 Db 181 LTLIDNTRSRACHPCSPMKGSRWGESSEDCOSLTRTVCAGGCARCKGLPTDCCHEQC 240
 QY 241 AAGCTGPKHSDCLACLFHNSHIGELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300
 Db 241 AAGCTGPKHSDCLACLFHNSHIGELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300
 QY 301 YNYLSTDVGSCTLVCPHNOEVTAEQGTORCEKSKPCARVCYGLGMOYIKANSKF 360
 Db 301 YNYLSTDVGSCTLVCPHNOEVTAEQGTORCEKSKPCARVCYGLGMOYIKANSKF 360
 QY 361 ELEFAGCKIFGSLAFPLPESFQDGPASNTAPLOPEQLQVFETLEEITGYLYISAWP 420
 Db 361 IQEFAGCKIIFGSLAFPLPESFQDGPASNTAPLOPEQLQVFETLEEITGYLYISAWP 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGSIWGLSLRELGLALIHNTLHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGSIWGLSLRELGLALIHNTLHLCFVHTV 480
 QY 481 PWDQLFRNPHQALLHTANRPEDECVCEGLACHQLCARGHCWGPGTQCVCNCSOFLRQEC 540
 Db 481 PWDQLFRNPHQALLHTANRPEDECVCEGLACHQLCARGHCWGPGTQCVCNCSOFLRQEC 540
 QY 541 VEBCRVQLGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVCAHYKDPFVCVARC 600
 Db 541 VEBCRVQLGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVCAHYKDPFVCVARC 600
 QY 601 PSGVKPDLSPYMPIWKFPEDEGACQPCPINCTHSCVDLDDKGCFAEQRASPLTSIVSAV 660
 Db 601 PSGVKPDLSPYMPIWKFPEDEGACQPCPINCTHSCVDLDDKGCFAEQRASPLTSIVSAV 660
 QY 661 ILLVVLGVVFGILIKRQOKIRKYTMRLLOBTTELVEPLTPSGAMPNOAOMRILKETEL 720
 Db 661 ILLVVLGVVFGILIKRQOKIRKYTMRLLOBTTELVEPLTPSGAMPNOAOMRILKETEL 720
 QY 721 RKVKVLGSGAFGQYIKA-----NSKFIGITELVRENTSPKANKEIIDEA VVMAGV 773
 Db 721 RKVKVLGSGAFGQYIKVIGIPDGENVK-IPVAIKVRENTSPKANKEIIDEA VVMAGV 779
 QY 774 PYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRNRRGLSGSQDLLNWCMTAKGMSYLEDV 833

Db 780 PYVSRLLGICLTSTVQLVTPMPYGCGLLDHVRNKRGLGSDLLNWCWQIAKMSYLEDV 839
Qy 834 RLVRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGGKVPKWALESILRRRF 893
Db 840 RLVRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGGKVPKWALESILRRRF 899
Qy 894 THQSDVMSYGVTVWELMTFAGKPDGIPAREIPDLLEKGERLPPOPICTIDVYMWKWCW 953
Db 900 THQSDVMSYGVTVWELMTFAGKPDGIPAREIPDLLEKGERLPPOPICTIDVYMWKWCW 959
Qy 954 MIDSECRPRFRELVSFMRMARDQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVD 1013
Db 960 MIDSECRPRFRELVSFMRMARDQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVD 1019
Qy 1014 ABEYLVPQGFCDPAPAGAGWVHRHRSSTSRSGGDLTLGLEPSEEEAPRSLAPSE 1073
Db 1020 ABEYLVPQGFCDPAPAGAGWVHRHRSSTSRSGGDLTLGLEPSEEEAPRSLAPSE 1079
Qy 1074 GAGSDVFDGDLGMAAGLQSLPHTDPSLQRYSEDPTVPLPSETDGVVAPLTCSPQPEY 1133
Db 1080 GAGSDVFDGDLGMAAGLQSLPHTDPSLQRYSEDPTVPLPSETDGVVAPLTCSPQPEY 1139
Qy 1134 VNQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVFAFGGAVENPEYLTLP 1193
Db 1140 VNQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVFAFGGAVENPEYLTLP 1199
Qy 1194 QGGAAPQHPPPAPSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGDLVVP 1249
Db 1200 QGGAAPQHPPPAPSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGDLVVP 1255

RESULT 5

AAW01111 standard; Protein; 1255 AA.
XX AC AAW01111;
DT 01-JAN-1997 (first entry)
XX DE HER-2/neu protein.
XX DE HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW breast cancer; ovary cancer; colon cancer; lung cancer;
KW prostate cancer; immunisation; tumour; vaccine; vector.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Domain 676..1255
FT /label= Intracellular domain
FT /note= "Claimed domain, useful for immunisation"

XX W09630514-AAI.

XX XX XX

XX 03-OCT-1996.

XX 28-MAR-1996; 96WO-US01689.

XX 31-MAR-1995; 95US-0414417.

XX (UNIW) UNIV WASHINGTON.

XX Cheever MA, Disis ML;

XX WPI; 1996-455361/45.

XX N-PSDB; AAT40739.

XX DNA encoding HER-2-neu poly-peptide(s) - used for prevention or

XX treatment of malignancies with which the HER-2/neu oncogene is

XX associated

XX Claim 2; Page 56-61; 71pp; English.

XX PS

XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.
XX SQ Sequence 1255 AA;
Query Match 97.4%; Score 6602; DB 17; Length 1255;
Best Local Similarity 97.2%; Pred No 0;
Matches 1221; Conservative 6; Mismatches 21; Indels 8; Gaps 2;
Qy 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLRLPASPEHLMLRHLHYQGCVVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLRLPASPEHLMLRHLHYQGCVVQGNL 60
Qy 61 ELYLPTNASLSFLQDIEQVGVVLAHNOVROVPLORLRIVRGTQLFEDNYALAVLDNG 120
Db 61 ELYLPTNASLSFLQDIEQVGVVLAHNOVROVPLORLRIVRGTQLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIORNPOLCYQDTILWKDIFHNQOLA 180
Db 121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIORNPOLCYQDTILWKDIFHNQOLA 180
Qy 181 LTLIDNRSRACHPCSPMCKGSRGWESSEDCQSLTRTVCAAGCARCKGLPTDCCHEQC 240
Db 181 LTLIDNRSRACHPCSPMCKGSRGWESSEDCQSLTRTVCAAGCARCKGLPTDCCHEQC 240
Qy 241 AAGCTGPKSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
Db 241 AAGCTGPKSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
Qy 301 YNYLTDVSGCTLVCPLNHNOVTAEDGTQCEKSCPKARVCYGLGMVYIKANSKFIGIT 360
Db 301 YNYLTDVSGCTLVCPLNHNOVTAEDGTQCEKSCPKARVCYGLGMVYIKANSKFIGIT 360
Qy 361 ELEFAGCKKIFGSLAFELPESFDGDPASNTAPLQEQVFPETLEEITGYLYISAWPDSLP 420
Db 361 IQEFAGCKKIFGSLAFELPESFDGDPASNTAPLQEQVFPETLEEITGYLYISAWPDSLP 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGSLRSLRGLSGLALIHNTLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGSLRSLRGLSGLALIHNTLHLCFVHTV 480
Qy 481 PWDQLFNPHOALLHTANRPEDECVGEGACHOLCARGHCWGPCTQVNCVSQFLRGQEC 540
Db 481 PWDQLFNPHOALLHTANRPEDECVGEGACHOLCARGHCWGPCTQVNCVSQFLRGQEC 540
Qy 541 VEESCRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600
Db 541 VEESCRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600
Qy 601 PSGVKPDLSPYMPIWKFPDEBACQPCPINCTHSCVDLDDKGCAPQASPLTISVAVG 660
Db 601 PSGVKPDLSPYMPIWKFPDEBACQPCPINCTHSCVDLDDKGCAPQASPLTISVAVG 660
Qy 661 ILLVVLGVVFGILIKRROQKIRKYMRLRLQTELVEPLTPSGAMPNQAMRLKETEL 720
Db 661 ILLVVLGVVFGILIKRROQKIRKYMRLRLQTELVEPLTPSGAMPNQAMRLKETEL 720
Qy 721 RKVKVLGSGAFGQYIKA-----NSKFIGITELVLENTSPKANKEILDEAYVMAGVGS 773
Db 721 RKVKVLGSGAFGQYIKA-----NSKFIGITELVLENTSPKANKEILDEAYVMAGVGS 779
Qy 774 PYVSRLLGICLTSTVQLVTPMPYGCGLLDHVRNKRGLGSDLLNWCWQIAKMSYLEDV 833
Db 780 PYVSRLLGICLTSTVQLVTPMPYGCGLLDHVRNKRGLGSDLLNWCWQIAKMSYLEDV 839

QY 834 RLVRDLAARNVLKSPNHHVKITDGLARLLDIDETEHADGGKVPKMALESILRRF 893
DB 840 RLVRDLAARNVLKSPNHHVKITDGLARLLDIDETEHADGGKVPKMALESILRRF 899
QY 894 THQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYMWKWCW 953
DB 900 THQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYMWKWCW 959
QY 954 MIDSECRPRFELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVD 1013
DB 960 MIDSECRPRFELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVD 1019
QY 1014 ABEYLVPQCGFFCDDPAPAGAGMVHHRSSSTSGGGDLTLGLEPSEERAPRSLAPSE 1073
DB 1020 ABEYLVPQCGFFCDDPAPAGAGMVHHRSSSTSGGGDLTLGLEPSEERAPRSLAPSE 1079
QY 1074 GAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEY 1133
DB 1080 GAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEY 1139
QY 1134 VNQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVAFPGGAVENPEYLT 1193
DB 1140 VNQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVAFPGGAVENPEYLT 1199
QY 1194 QGGAPOPHPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGDLVVPV 1249
DB 1200 QGGAPOPHPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGDLVVPV 1255

RESULT 6

AAW92406
ID AAW92406 standard; Protein; 1255 AA.
XX
AC AAW92406;
XX AC
XX 21-APR-1999 (first entry)
XX
DE Human HER-2/neu oncogene protein.
XX
XX HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
KW malignancy; treatment; tumour.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH Region 676..1255
FT /note="region which elicits immune response"
XX
XX US5869445-A.
XX
XX 09-FEB-1999.
XX
XX 01-APR-1996; 96US-0625101.
XX
XX 01-APR-1996; 96US-0625101.
PR 17-MAR-1993; 93US-0033644.
PR 12-AUG-1993; 93US-0106112.
PR 31-MAR-1995; 95US-0414417.
XX
XX (UNIW) UNIV WASHINGTON.
XX
XX Cheever MA, Disis ML;
XX WPI; 1999-152835/13.
DR N-PSDB; AAX01912.
XX

XX Use of HER-2/neu polypeptides - for eliciting an immune response to
PT an HER-2/neu associated malignancy, particularly for treating or
PT preventing tumours
XX

PS Claim 3; Column 31-38; 26pp; English.
XX

XX This sequence represents the human HER-2/neu oncogene protein. A fragment
CC

CC of this protein is used in a method for eliciting or enhancing an immune
CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
CC B cells to produce an immune response to the HER-2/neu protein. The
CC method can be used for immunisation against a malignancy in which the
CC HER-2/neu oncogene is associated and in the treatment of an existing
CC tumour, or to prevent tumour occurrence or reoccurrence.
XX

SQ Sequence 1255 AA;

Query Match 97.4%; Score 6602; DB 20; Length 1255;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 1221; Conservative 6; Mismatches 21; Indels 8; Gaps 2;

QY 1 MELAAALCRWCLLALLPPGAASQVCTGTDKMLRLPASPTHLDMLRHLHYGGQVWGQNL 60
DB 1 MELAAALCRWCLLALLPPGAASQVCTGTDKMLRLPASPTHLDMLRHLHYGGQVWGQNL 60
QY 61 ELTYLPTNALSFLQDIQEVQGVLTAAHNOVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
DB 61 ELTYLPTNALSFLQDIQEVQGVLTAAHNOVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
QY 121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQENPOLCYQDTILWKDIFHKNOLA 180
DB 121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQENPOLCYQDTILWKDIFHKNOLA 180
QY 181 LTLIDNRSRACHPCSPMKGSCWGESSEDCQSLTRTVCCAGCARCKGPLDCCHEQC 240
DB 181 LTLIDNRSRACHPCSPMKGSCWGESSEDCQSLTRTVCCAGCARCKGPLDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVSGCTLVCPLNHOEVTAEQTCRCEKSPCARVCYGLGMQYIKANSKFIGIT 360
DB 301 YNYLSTDVSGCTLVCPLNHOEVTAEQTCRCEKSPCARVCYGLGMQYIKANSKFIGIT 360
QY 361 ELEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOEQLOVFETLEEITGYLYISAWPDSLP 420
DB 361 IQEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOEQLOVFETLEEITGYLYISAWPDSLP 420
QY 421 DLSVFQNLQVIRGRIILHNGAYSILTQGLGISWLGSLRSLRGLSGLALIHNTLHLCFVHTV 480
DB 421 DLSVFQNLQVIRGRIILHNGAYSILTQGLGISWLGSLRSLRGLSGLALIHNTLHLCFVHTV 480
QY 481 PWDLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGPTQVCNCSQFLRGQEC 540
DB 481 PWDLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGPTQVCNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREYVVARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFVCVARC 600
DB 541 VEECRVLQGLPREYVVARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFVCVARC 600
QY 601 PSGVKPDLSPYMPIWKPDPBEGACQPCINCTHSCVDLDDKGCBAEQRASPLTSIVSAVG 660
DB 601 PSGVKPDLSPYMPIWKPDPBEGACQPCINCTHSCVDLDDKGCBAEQRASPLTSIVSAVG 660
QY 661 ILLVVVLGVVFGTILIKRROOKIRKYTMRLRLLQSTELVEPLTPSGAMPNQAMRLKETEL 720
DB 661 ILLVVVLGVVFGTILIKRROOKIRKYTMRLRLLQSTELVEPLTPSGAMPNQAMRLKETEL 720
QY 721 RKVKVLGSGAFGQYIKA-----NSKFIGITELVLRNTSPKANKEILDEAYVMAGVGS 773
DB 721 RKVKVLGSGAFGTVYKGIWIPDGENVK-IPVALKVLRENTSPKANKEILDEAYVMAGVGS 779
QY 774 PYVSRLLGICLTSTVOLVTQLMYPYGLLDHVRNRRGLSGQDLLNWCMIKAGMSYLEDV 833
DB 780 PYVSRLLGICLTSTVOLVTQLMYPYGLLDHVRNRRGLSGQDLLNWCMIKAGMSYLEDV 839
QY 834 RLVRDLAARNVLKSPNHHVKITDGLARLLDIDETEHADGGKVPKMALESILRRF 893
DB 840 RLVRDLAARNVLKSPNHHVKITDGLARLLDIDETEHADGGKVPKMALESILRRF 899

Qy 894 THQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYIMVKWCW 953
Dy 900 THQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYIMVKWCW 959
Qy 954 MIDSECRPRFRELVSERSMARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMGDLVD 1013
Dy 960 MIDSECRPRFRELVSERSMARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMGDLVD 1019
Qy 1014 ABEYLVPQGFCDPAPAGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSE 1073
Dy 1020 ABEYLVPQGFCDPAPAGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSE 1079
Qy 1074 GAGSDVFDGDLGMAAGLQSLTHDRSPLOQRYSEDPTVPLPSETDGYVAPLTCSPOPEY 1133
Dy 1080 GAGSDVFDGDLGMAAGLQSLTHDRSPLOQRYSEDPTVPLPSETDGYVAPLTCSPOPEY 1139
Qy 1134 VNQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLT 1193
Dy 1140 VNQDVRPQPPSPREGPLPAARPAAGATLERPKTILSPGKNGVVKOVFAFGGAVENPEYLT 1199
Qy 1194 QGGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGPTAENPEYLGDLV 1249
Dy 1200 QGGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGPTAENPEYLGDLV 1255

RESULT 7

ID AAB21198 standard; protein; 1255 AA.

AC AAB21198;

DT 12-JAN-2001 (first entry)

DE Human HER-2/neu protein.

KW Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
KW colon cancer.

OS Homo sapiens.

PN WO200044899-A1.

PD 03-AUG-2000.

PF 28-JAN-2000; 2000WO-US02164.

PR 29-JAN-1999; 99US-0117976.

PA (CORI-) CORIXA CORP.

PA (SMIK) SMITHKLINE BEECHAM.

PI Cheever MA, Gheysen D;

XX WPI; 2000-505976/45.

DR N-PSDB; AAA89736.

PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT useful for vaccinating against breast, ovarian, colon, lung and
PT prostate cancers -

PS Claim 52; Fig 7; 128pp; English.

CC The present sequence is the human HER-2/neu protein. It is a member of
CC the tyrosine kinase family of receptor-like glycoproteins and shows
CC homology to the epidermal growth factor receptor (EGFR). It probably
CC plays a part in cell growth and/or differentiation. The HER-2/neu
CC gene is an oncogene. An HER-2/neu fusion protein comprising a
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC domain may be used to treat or prevent cancer by eliciting or
CC enhancing an immune response to the HER-2/neu protein. It may be used
CC to treat malignancies such as breast, ovarian, colon, lung and
CC prostate cancers, and may be used as an antigen to vaccinate againstCC these neoplasias.
XX
SQ Sequence 1255 AA;Query Match 97.4%; Score 6602; DB 21; Length 1255;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 1221; Conservative 6; Mismatches 21; Indels 8; Gaps 2;

Qy 1 MELAALCRWGLLALLPPGAASQVCTGTDMKLRLPASPTHLDMLRHLHQGCVVQGNL 60
Dy 1 MELAALCRWGLLALLPPGAASQVCTGTDMKLRLPASPTHLDMLRHLHQGCVVQGNL 60
Qy 61 ELTYLPTNASLSLFDIQEYVQGVLIHAHQVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
Dy 61 ELTYLPTNASLSLFDIQEYVQGVLIHAHQVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNOLA 180
Dy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNOLA 180
Qy 181 LTLIDTNRSRACHPCSPMCKSGSCWGESSEDCSLTRTVCCAGCARCKGPLPTDCCHEQC 240
Dy 181 LTLIDTNRSRACHPCSPMCKSGSCWGESSEDCSLTRTVCCAGCARCKGPLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Dy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVSGCTLVCPLNHOEVTAEDGTQRCCKSPCARVCYGLGHOYIKANSKFTGIT 360
Dy 301 YNYLSTDVSGCTLVCPLNHOEVTAEDGTQRCCKSPCARVCYGLGHOYIKANSKFTGIT 360
Qy 361 ELEFAGCKIFGSLAFIPESFDGDPASNTAPLOEQVETLEEITGYLYISAWPDSLP 420
Dy 361 IQBFAGCKIFGSLAFIPESFDGDPASNTAPLOEQVETLEEITGYLYISAWPDSLP 420
Qy 421 DLSVFQNLQVIRGRIHLHGAYSLTLOGLIGISWLSRLSRLSGLALIHNTLHLCFVHTV 480
Dy 421 DLSVFQNLQVIRGRIHLHGAYSLTLOGLIGISWLSRLSRLSGLALIHNTLHLCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQI CARGHCWGPGPTQVNCSPQLRQEC 540
Dy 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQI CARGHCWGPGPTQVNCSPQLRQEC 540
Qy 541 VESCRVLQGLPREVYNARHCLPCHPECPONGSVTCFPEADOCVACAHYKDPFPFCVARC 600
Dy 541 VESCRVLQGLPREVYNARHCLPCHPECPONGSVTCFPEADOCVACAHYKDPFPFCVARC 600
Qy 601 PSGVKPDLSTYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQRASPLTSISAVVG 660
Dy 601 PSGVKPDLSTYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQRASPLTSISAVVG 660
Qy 661 ILLVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Dy 661 ILLVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLGSGAFGOYIKA-----NSKFIGITELVIRENTSPKANKEILDEAYVMAGVGS 773
Dy 721 RKVKVLGSGAFGOYIKGIWIPDGENVK-IPVALKVLRENTSPKANKEILDEAYVMAGVGS 779
Qy 774 PYVSRLLIGICLTSTVQLVTQLMFYGCLLDHVRNRRGLSGDQLLNWCQIAKMSYLEDV 833
Dy 780 PYVSRLLIGICLTSTVQLVTQLMFYGCLLDHVRNRRGLSGDQLLNWCQIAKMSYLEDV 839
Qy 834 RLVRDLAARNVLKSPNHHKVTDFGLARLLDIDETEHADGGKVPKKNWALESIILRRRF 893
Dy 840 RLVRDLAARNVLKSPNHHKVTDFGLARLLDIDETEHADGGKVPKKNWALESIILRRRF 899
Qy 894 THQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYIMVKWCW 953
Dy 900 THQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYIMVKWCW 959
Qy 954 MIDSECRPRFRELVSERSMARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMGDLVD 1013

|||||
Db 960 MIDSECRPRFELVSEFSRWARDPQRFVVIQNEGLGASPLDSTFYRSLLEDDMGDLVD 1019
Qy 1014 AEEYLVPQGFCDPAPAGAGMHHRRSSSTRSGGDLTLGLEPSEEEAPRSPAPSE 1073
Db 1020 AEEYLVPQGFCDPAPAGAGMHHRRSSSTRSGGDLTLGLEPSEEEAPRSPAPSE 1079
Qy 1074 GAGSDVFDGLGMAAGLQSLPHDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEY 1133
Db 1080 GAGSDVFDGLGMAAGLQSLPHDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEY 1139
Qy 1134 VNQDVRPQPPSPREGPLPAARPAATLERAKTLSPGKNGVVKDVFAGGAVENPEYLT 1193
Db 1140 VNQDVRPQPPSPREGPLPAARPAATLERAKTLSPGKNGVVKDVFAGGAVENPEYLT 1199
Qy 1194 QGGAAPQHPHPPAPAFDNLYYWDQPPPERGAPPSTFKGPTTAENPEYLGDLVPV 1249
Db 1200 QGGAAPQHPHPPAPAFDNLYYWDQPPPERGAPPSTFKGPTTAENPEYLGDLVPV 1255

RESULT 8

AAY84780
ID AAY84780 standard; Protein; 1255 AA.

XX AC AAY84780;

XX DT 08-AUG-2000 (first entry)

XX DE Amino acid sequence of the SPLICE erbb-2 receptor protein.

XX KW SPLICE erbb-2 receptor protein; cell transformation disorder; cancer;
KW tumor cell proliferation; tissue degeneration; arthropathy;
KW bone resorption; inflammatory disease; degenerative disorder;
KW wound healing.

XX OS Homo sapiens.

XX PN W0200020579-A1.

XX PD 13-APR-2000.

XX PF 01-OCT-1999; 99WO-CA00912.

XX PR 02-OCT-1998; 98US-0165192.

XX PA (UYMC-) UNIV MCMASTER.

XX PI Muller WJ, Siegel PM;

XX WPI; 2000-303768/26.

XX DR N-PSDB; AAA14812.

XX PT Nucleic acid encoding an erbb 2 receptor protein designated SPLICE
PT erbb-2, inhibitors of the protein are useful for treatment of cancer -
XX PS Claim 3; Fig 2; 60pp; English.

CC The present sequence represents a SPLICE erbb-2 receptor protein. The
CC protein has an in-frame deletion of 16 amino acids, 2 of which are
CC conserved cysteine residues, compared to the unspliced protein. The
CC erbb-2 polynucleotide is used to construct probes for detecting
CC disorders of cell transformation such as cancer. Antibodies to the
CC protein may be used to detect SPLICE erbb-2 in a sample. Agents
CC (e.g. antisense oligonucleotides) which inhibit the expression of
CC SPLICE erbb-2 are useful for reducing tumor cell proliferation and
CC treating cancer. Substances which stimulate SPLICE erbb-2 are useful
CC for treating conditions involving damaged cells including conditions
CC in which degeneration of tissue occurs, such as arthropathy, bone
CC resorption, inflammatory diseases, degenerative disorders of the
CC central nervous system and wound healing.

XX Sequence 1255 AA;

Qy 1014 AEEYLVPQGFCDPAPAGAGMHHRRSSSTRSGGDLTLGLEPSEEEAPRSPAPSE 1073

Query Match 97.4%; Score 6602; DB 21; Length 1255;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 1221; Conservative 6; Mismatches 21; Indels 8; Gaps 2;
Qy 1 MEALAAACRWGLLALLPPGAASQVCTGTDMKRLRLPASPTHLDMLRHLHLYOGCOVVOGNL 60
Db 1 MEALAAACRWGLLALLPPGAASQVCTGTDMKRLRLPASPTHLDMLRHLHLYOGCOVVOGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQGVYLAHNOVROVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQGVYLAHNOVROVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Qy 121 DPLNTPPTVGTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTTLWKDIFHNNOLA 180
Db 121 DPLNTPPTVGTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTTLWKDIFHNNOLA 180
Qy 181 LTLIDTNRSEACHPCSPMKSGSCWGESSEDCOSLRTVTCAGGCARCKGPLPTDCCHEQC 240
Db 181 LTLIDTNRSEACHPCSPMKSGSCWGESSEDCOSLRTVTCAGGCARCKGPLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLHFNHSGICEHLCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLHFNHSGICEHLCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPLNHQEVTAEDGTQRCCKSKPCARVCYGLGMHLEHRAVTSAN 360
Db 301 YNYLSTDVGSCTLVCPLNHQEVTAEDGTQRCCKSKPCARVCYGLGMHLEHRAVTSAN 360
Qy 361 ELSEFACKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFEETLEEITGYLYISAWPSLP 420
Db 361 IQBFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFEETLEEITGYLYISAWPSLP 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLLTLOGLGISWLGSLRSLRGLSGLALIHNTLHCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLLTLOGLGISWLGSLRSLRGLSGLALIHNTLHCFVHTV 480
Qy 481 PWDOLFNRPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQCVNCSQFLRQEC 540
Db 481 PWDOLFNRPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQCVNCSQFLRQEC 540
Qy 541 VEESRVLOGLPREYVNAHCLPCHPECPONGSVTCFEPADQCVACAHYKDPFCVARC 600
Db 541 VEESRVLOGLPREYVNAHCLPCHPECPONGSVTCFEPADQCVACAHYKDPFCVARC 600
Qy 601 PSGVKPDLSPYMPIWKPDEBAGQPCPINCHSCVDLDDKCGCAEORASPLTSTVSAVVG 660
Db 601 PSGVKPDLSPYMPIWKPDEBAGQPCPINCHSCVDLDDKCGCAEORASPLTSTVSAVVG 660
Qy 661 ILLVWVGVVFGILIKRQOKIRKYTMRLLOQETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVWVGVVFGILIKRQOKIRKYTMRLLOQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLGSGAFQYIKA-----NSKFGITELVIRENTSPKANKEILDEAYVMAGVGS 773
Db 721 RKVKVLGSGAFQYIKGWIWIPDGENVK-IPVAIKVIRENTSPKANKEILDEAYVMAGVGS 779
Qy 774 PYVSRLLGICLTSTVOLVTOLMPYGLLDHVRNRRGLSGSODLLNMCQIAKMSYLEDV 833
Db 780 PYVSRLLGICLTSTVOLVTOLMPYGLLDHVRNRRGLSGSODLLNMCQIAKMSYLEDV 839
Qy 834 RLVRDLAARNVLKSPNVHKITDFGLARLLDDIDETEHADGKGVPIKWMALRESILRRRF 893
Db 840 RLVRDLAARNVLKSPNVHKITDFGLARLLDDIDETEHADGKGVPIKWMALRESILRRRF 899
Qy 894 THQSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPPQPICTIDVTMVMVKCW 953
Db 900 THQSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPPQPICTIDVTMVMVKCW 959
Qy 954 MIDSECRPRFELVSEFSRWARDPQRFVVIQNEGLGASPLDSTFYRSLLEDDMGDLVD 1013
Db 960 MIDSECRPRFELVSEFSRWARDPQRFVVIQNEGLGASPLDSTFYRSLLEDDMGDLVD 1019
Qy 1014 AEEYLVPQGFCDPAPAGAGMHHRRSSSTRSGGDLTLGLEPSEEEAPRSPAPSE 1073

Db 1140 VNQDVRPQPPSPREGELPAARPAAGATLERPKTSLSPKNGVVKDVFAFGGAVENPEYLTP 1199
Qy 1194 QGGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGDLVDPV 1249
Db 1200 QGGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGDLVDPV 1255
RESULT 10
AAG88267
ID AAG88267 standard; Protein; 1255 AA.
AC AAG88267;
XX
XX 11-SEP-2001 (first entry)
XX
DE HER2/neu amino acid sequence.
XX
XX Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
XX Homo sapiens.
XX
XX WO200141787-A1.
XX
XX 14-JUN-2001.
XX
XX 11-DEC-2000; 2000WO-US33591.
XX
XX 10-DEC-1999; 99US-0458299.
XX
XX (EPIM-) EPIMMUNE INC.
XX
XX Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
PI Keogh E;
XX
XX WPI; 2001-374995/39.
XX
XX An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer -
XX
XX Disclosure; Page 15; 1999p; English.
XX
XX The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (1)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
CC and immunostimulant activities, and can be used in vaccines. (I), (II)
CC and (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample form a patient and
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour
CC tumour variability and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention.
XX
XX Sequence 1255 AA;

Query Match

97.4%; Score 6602; DB 22; Length 1255;

Db 1020 ABEYLVPQQGFFCPDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSE 1079
 QY 1074 GAGSDVFDGLGMAAGLQSLPTHDSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEY 1133
 Db 1080 GAGSDVFDGLGMAAGLQSLPTHDSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEY 1139
 QY 1134 VNQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVVDVFAFGGAVENPEYLTP 1193
 Db 1140 VNQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVVDVFAFGGAVENPEYLTP 1199
 QY 1194 QGGAAPHPHPPAPSPAFDNLVYWDQDPPERGAPSTFKGPTTAENPEYGLDVPV 1249
 Db 1200 QGGAAPHPHPPAPSPAFDNLVYWDQDPPERGAPSTFKGPTTAENPEYGLDVPV 1255

RESULT 11

AAE24067
 ID AAE24067 standard; Protein; 1255 AA.

AC AAE24067;

DT 23-SEP-2002 (first entry)

XX Human Her-2 protein.

XX Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
 KW hyperproliferative disorder; prophylaxis; inflammation; antisense;
 KW tumour; gene therapy; phosphorothioate backbone.

XX Homo sapiens.

XX WO200222636-A1.

XX 21-MAR-2002.

XX 12-SEP-2001; 2001WO-US28572.

XX 15-SEP-2000; 2000US-0663834.

XX (ISIS-) ISIS PHARM INC.

XX Bennett CF, Cowser LM;

XX WPI; 2002-471192/50.

XX N-PSDB; RAD38904.

XX Novel antisense oligonucleotide which modulates the expression of Human
 PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
 PT inflammation or to prevent infection in humans -

XX Example 13; Page 95-107; 116pp; English.

XX The invention relates to antisense compounds targetted to a nucleic
 CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
 CC that specifically hybridises with and inhibits the expression of Her2.
 CC Antisense compounds of the invention are used for treating diseases or
 CC conditions associated with Her2 such as hyperproliferative disorders
 CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
 CC neural or cardiac cancer. They are also useful prophylactically e.g.
 CC to prevent or delay infection, inflammation and tumour formation. The
 CC invention is also used in gene therapy. The present sequence is human
 CC Her-2 protein.

XX Sequence 1255 AA;

Query Match 97.4%; Score 6602; DB 23; Length 1255;

Best Local Similarity 97.2%; Pred. No. 0;

Matches 1221; Conservative 6; Mismatches 21; Indels 8; Gaps 2;

QY 1 MELAALCRWGLLALLPPGAASCTGCTDKMLRLPASPETHLDMRLHYGGCCVVOGNL 60

Db 1 MELAALCRWGLLALLPPGAASCTGCTDKMLRLPASPETHLDMRLHYGGCCVVOGNL 60

QY 61 ELTYLPTNASLSLQDIQEVQGVYLAHQVQVPLQRLRIVRGTQOLFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLSLQDIQEVQGVYLAHQVQVPLQRLRIVRGTQOLFEDNYALAVLDNG 120
 QY 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNOLA 180
 Db 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNOLA 180
 QY 181 LTLIDNTRSPACHPCSPMKCSGSESSDCSLTRTVCCAGCACRCKGPLPTDCCHEQC 240
 Db 181 LTLIDNTRSPACHPCSPMKCSGSESSDCSLTRTVCCAGCACRCKGPLPTDCCHEQC 240
 QY 241 AAGCTGPKSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACP 300
 Db 241 AAGCTGPKSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACP 300
 QY 301 YNYLSTDVSGCTLVCPLNHQEVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKFTGIT 360
 Db 301 YNYLSTDVSGCTLVCPLNHQEVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKFTGIT 360
 QY 361 ELSPAGCKKIFGSLAFLPESFDGDPASNTAPLQEQLOVPETLEEITGYLYISAWPDSL 420
 Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQEQLOVPETLEEITGYLYISAWPDSL 420
 QY 421 DLSVFQNLQVIRGRIILHNGAYSLTLOGLGISWLSRLSRLSGLALIHNTLHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRIILHNGAYSLTLOGLGISWLSRLSRLSGLALIHNTLHLCFVHTV 480
 QY 481 PWDOLFNRPHQALLHTANRPEDCEVGEGLACHQLCARGHCWGPGTQCVNCSQFLRQEC 540
 Db 481 PWDOLFNRPHQALLHTANRPEDCEVGEGLACHQLCARGHCWGPGTQCVNCSQFLRQEC 540
 QY 541 VEECRVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFCVARC 600
 Db 541 VEECRVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFCVARC 600
 QY 601 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCFAEORASPLTSTVAVG 660
 Db 601 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCFAEORASPLTSTVAVG 660
 QY 661 ILLVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
 Db 661 ILLVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
 QY 721 RKVKVLGSGAFGOYIKA-----NSKFTGITELVRENTSPKANKEILDEAYVMAGVS 773
 Db 721 RKVKVLGSGAFGOYIKAIIPDGENVK-IPVAIKVRENTSPKANKEILDEAYVMAGVS 779
 QY 774 PYVSRLLIGICLTSTVQLVTQLMFYGCLLDHVRENRLGSLQDILLNMCQIAKMSYLEDV 833
 Db 780 PYVSRLLIGICLTSTVQLVTQLMFYGCLLDHVRENRLGSLQDILLNMCQIAKMSYLEDV 839
 QY 834 RLVRDLAARNVLVKSNNHVKITDFGLARLLDIDETEHADGKVPKIMMALESILRRRF 893
 Db 840 RLVRDLAARNVLVKSNNHVKITDFGLARLLDIDETEHADGKVPKIMMALESILRRRF 899
 QY 894 THQSDVMSYGVTVWELMTGAKYDGIIPAREIPDLLEKGERLPQPPCTTIDVVMVWKW 953
 Db 900 THQSDVMSYGVTVWELMTGAKYDGIIPAREIPDLLEKGERLPQPPCTTIDVVMVWKW 959
 QY 954 MIDSECRPRELVSFSESRMARDPQRFVVIQNEDELGPASPLDSTFYRSLLEDDMDGLVD 1013
 Db 960 MIDSECRPRELVSFSESRMARDPQRFVVIQNEDELGPASPLDSTFYRSLLEDDMDGLVD 1019
 QY 1014 ABEYLVPQQGFFCPDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSE 1073
 Db 1020 ABEYLVPQQGFFCPDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSE 1079
 QY 1074 GAGSDVFDGLGMAAGLQSLPTHDSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEY 1133
 Db 1080 GAGSDVFDGLGMAAGLQSLPTHDSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEY 1139
 QY 1134 VNQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVVDVFAFGGAVENPEYLTP 1193

||||| 1140 VNQDVRPQPPSPREGELPAARPAATLERPKTUSPGKNGVVDVAFPGGAVENPEYLTP 1199
1194 QGGAAPQHPHPPAPAFDNLYYWDQPPPERGAPPSTFKGTPTAENPEYGLDVPV 1249
1200 QGGAAPQHPHPPAPAFDNLYYWDQPPPERGAPPSTFKGTPTAENPEYGLDVPV 1255

RESULT 12
AAE20479
ID AAE20479 standard; Protein; 1255 AA.
XX AC AAE20479;
XX DT 01-JUL-2002 (first entry)
XX DE Human Her-2/neu protein.
XX KW Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
XX KW human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX OS Homo sapiens.
XX PH Location/Qualifiers
XX FT 1021..1030
XX FT /note= "Naturally processed HLA-B44-restricted epitope"
XX PN WO200214503-A2.
XX PD 21-FEB-2002.
XX PF 14-AUG-2001; 2001WO-US41733.
XX PR 14-AUG-2000; 2000US-225152P.
XX PR 28-SEP-2000; 2000US-236428P.
XX PR 21-FEB-2001; 2001US-270520P.
XX PA (CORI-) CORIXA CORP.
XX PI Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
PI Mcneill PD, Vedvick TS;
XX WI N-PSDB; AAD32743.
XX DR WPI; 2002-280758/32.
XX PT Novel isolated Her-2/Neu polypeptide composition useful for therapy,
XX PT prevention and diagnosis of cancer, preferably breast cancer -
XX PS Disclosure; Page 114-117; 129pp; English.
XX CC The invention relates to an isolated Her-2/Neu polypeptide composition
XX CC effective for eliciting an immune response. The invention is useful for
XX CC eliciting an immune response in a patient, where the patient is human
XX CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
XX CC The composition is useful for the therapy and diagnosis of cancer,
XX CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
XX CC and other compositions for the diagnosis, prevention and treatment of
XX CC human malignancies, for stimulating and/or expanding T cells specific for
XX CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a
XX CC patient. The invention is useful for stimulating a T cell response in a
XX CC human patient, as probe or primer for nucleic acid hybridisation, to
XX CC selectively form duplex molecules with complementary stretches of the
XX CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full
XX CC length gene from a suitable library, and to direct expression of a
XX CC polypeptide in appropriate host cells. The composition is useful in
XX CC prophylactic or therapeutic applications and for the treatment of cancer,
XX CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-
XX CC associated malignancies. The invention is useful in gene therapy. The
XX CC present sequence is human Her-2/neu protein.
XX Sequence 1255 AA;

Query Match 97.4%; Score 6602; DB 23; Length 1255;

Best Local Similarity 97.2%; Pred. No. 0; Matches 1221; Conservative 6; Mismatches 21; Indels 8; Gaps 2;	
Qy	1 MELAALCRWGLLLALLPPGAASQVCTGDMKLRLLPASPETHDMLRHLHLYGQGVVGNL 60
Db	1 MELAALCRWGLLLALLPPGAASQVCTGDMKLRLLPASPETHDMLRHLHLYGQGVVGNL 60
Qy	61 ELTYLPTNASLSFLQDIEVQGYVLIHANOVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Db	61 ELTYLPTNASLSFLQDIEVQGYVLIHANOVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Qy	121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQBNPOLCYQDTILWKDIFHKNOLA 180
Db	121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQBNPOLCYQDTILWKDIFHKNOLA 180
Qy	181 LTLIDTNRSRACHPCSPMKSGSCWGESSEDCOSLRTVTCAGGCARCKGLPTDDCCEQC 240
Db	181 LTLIDTNRSRACHPCSPMKSGSCWGESSEDCOSLRTVTCAGGCARCKGLPTDDCCEQC 240
Qy	241 AAGCTGPKGSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300
Db	241 AAGCTGPKGSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300
Qy	301 YNYLSTDVGSCTLVCPLNHOEVTABGTQRCCKSPCARVCYGLGMOYIKANSKFTGIT 360
Db	301 YNYLSTDVGSCTLVCPLNHOEVTABGTQRCCKSPCARVCYGLGMOYIKANSKFTGIT 360
Qy	361 ELEFAGCKIFGSLAFLPESFDGDPASNTAPLOEQLOVFEETLEETGYLYISAWPDSLP 420
Db	361 IQEFAGCKIFGSLAFLPESFDGDPASNTAPLOEQLOVFEETLEETGYLYISAWPDSLP 420
Qy	421 DLSVFQNLQVIRGRIILHNGAYSLTQGLGSIWGLSLRSLRGLSGLALIHNTLHCFVHTV 480
Db	421 DLSVFQNLQVIRGRIILHNGAYSLTQGLGSIWGLSLRSLRGLSGLALIHNTLHCFVHTV 480
Qy	481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCHWGPGTQCVNCSQFLRQEC 540
Db	481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCHWGPGTQCVNCSQFLRQEC 540
Qy	541 VEESRVLQGLPREYVVARHCLPCHPCOPONGSVTCFGEADOCVACAHYKDPFCVCARC 600
Db	541 VEESRVLQGLPREYVVARHCLPCHPCOPONGSVTCFGEADOCVACAHYKDPFCVCARC 600
Qy	601 PSGVKPDLSPYMKFPDEBEGACQPCINCTHSCVDLDDKGCPAEORASPLTISVAVVG 660
Db	601 PSGVKPDLSPYMKFPDEBEGACQPCINCTHSCVDLDDKGCPAEORASPLTISVAVVG 660
Qy	661 ILLVVVLGVVFGILIKRROQKIRKYMRLLOTELVEPLTPSGAMPNQAOMRILKETEL 720
Db	661 ILLVVVLGVVFGILIKRROQKIRKYMRLLOTELVEPLTPSGAMPNQAOMRILKETEL 720
Qy	721 RKVKVLGSGAFQGYIKA-----NSKFGITELVRENTSPKANKEILDEAYVMAGVGS 773
Db	721 RKVKVLGSGAFQGYIKVIMPDGENVK-IPVAIKVIRENTSPKANKEILDEAYVMAGVGS 779
Qy	774 PYVSRLLGLCTSTVOLVQLMPYGLLDHVRNRRGLSGODLLNWCMAQKAGMSYLEDV 833
Db	780 PYVSRLLGLCTSTVOLVQLMPYGLLDHVRNRRGLSGODLLNWCMAQKAGMSYLEDV 839
Qy	834 RLVRDLAARNVLKSPNVKHTDFGLARLLDIDETEHADGKVPKKNWMALESILRRRF 893
Db	840 RLVRDLAARNVLKSPNVKHTDFGLARLLDIDETEHADGKVPKKNWMALESILRRRF 899
Qy	894 THQSDVMSVGVTVWELMTGAKYDGIIPAREIPDLLEKGERLPQPPICITDVTVMVKCW 953
Db	900 THQSDVMSVGVTVWELMTGAKYDGIIPAREIPDLLEKGERLPQPPICITDVTVMVKCW 959
Qy	954 MIDSECRPRFRELVSFSSWARDPQRFVVTQNEIDLGPASPLDSTFYRSILLEDDMDGLVD 1013
Db	960 MIDSECRPRFRELVSFSSWARDPQRFVVTQNEIDLGPASPLDSTFYRSILLEDDMDGLVD 1019
Qy	1014 ABEYLVPQGGFFCPDPAPGAGGMVHRHRSSTSRGGDLTLGLEPSEEAEPRLAPSE 1073

Db 1020 AEEYLVPOQGFCDPAPGAGWVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSE 1079
Qy 1074 GAGSDVFDGLGMAAGLQSLPTHDPSPLOQYSEDPTVPLPSTGTVVAPLTCSPQPEY 1133
Db 1080 GAGSDVFDGLGMAAGLQSLPTHDPSPLOQYSEDPTVPLPSTGTVVAPLTCSPQPEY 1139
Qy 1134 VNQPDVFPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYLTP 1193
Db 1140 VNQPDVFPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYLTP 1199
Qy 1194 QGGAAPQHPHPPAPSPAFDNLVYWDQPPERGAPSPFTKGTPTAENPEYGLDVPV 1249
Db 1200 QGGAAPQHPHPPAPSPAFDNLVYWDQPPERGAPSPFTKGTPTAENPEYGLDVPV 1255

RESULT 13

AAM51143
ID AAM51143 standard; Protein; 1255 AA.

XX AC AAM51143;

XX DT 17-JUN-2002 (first entry)

XX DE Human Her-2/neu oncogene-encoded p185 glycoprotein.

XX KW Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;

XX KW tyrosine kinase; receptor; c-erbB2; gene therapy.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Domain 1..653 /note= "extracellular domain"

FT Domain 676..1255 /note= "intracellular domain"

FT Domain 990..1255 /note= "phosphorylation domain"

XX WO200212341-A2.

XX PN 14-FEB-2002.

XX PF 03-AUG-2001; 2001WO-US24283.

XX PR 03-AUG-2000; 2000US-0632507.

XX PA (CORI-) CORIXA CORP.

XX PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX PI Cheever MA, Gheysen D;

XX DR WPI; 2002-241743/29.

XX DR N-PSDB; ABA92250.

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting
or enhancing an immune response to the protein, has Her-2/neu
extracellular domain fused to Her-2/neu intracellular or
phosphorylation domain

XX Claim 68; Fig 7; 141pp; English.

XX The present sequence is that of human Her-2/neu (p185 glycoprotein

or c-erbB2), an oncogenic self-protein and target for anti-cancer

vaccines. The Her-2/neu gene is amplified and p185 is overexpressed

in a variety of cancers, including breast, ovarian, colon, lung and

prostate cancer. Her-2/neu is a member of the tyrosine kinase

family of receptor-like glycoproteins. It comprises an extracellular

domain with homology to the epidermal growth factor receptor

(EGFR), a highly hydrophobic transmembrane domain and a C-terminal

intracellular domain that also shows homology to EGFR. Its

overexpression correlates with a poor prognosis in breast and

ovarian cancers. The invention provides Her-2/neu fusion

proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In
CC preferred fusion proteins, the extracellular domain of a Her-2/neu
CC protein is fused to a Her-2/neu intracellular domain or
CC phosphorylation domain (or its DeltaPD fragment). An immune
CC response to Her-2/neu protein is elicited or enhanced by
CC administering the fusion protein in the form of a vaccine, or by
CC transfecting cells of an animal ex vivo with a nucleic acid
CC encoding the fusion protein, and delivering the transfected cells
CC to the animal. The fusion proteins, nucleic acids, and isolated
CC specific T-cells are useful for inhibiting the development of a
CC cancer, especially breast, ovarian, colon, lung or prostate cancer
CC in a patient. T cells that specifically react with a Her-2/neu
CC fusion protein can be used to remove tumour cells from a sample in
CC order to inhibit the development of cancer in a patient.

XX SQ Sequence 1255 AA;

Query Match 97.4%; Score 6602; DB 23; Length 1255;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 1221; Conservative 6; Mismatches 21; Indels 8; Gaps 2;

Qy 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKRLRPASPETHLDMRLHLYQGCVVQGNL 60

Db 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKRLRPASPETHLDMRLHLYQGCVVQGNL 60

Qy 61 ELTYLPTNASLSFLQDIQEVQGVVLAHNOVQVPLQRLRVGTQTFEDNYALAVLDNG 120

Db 61 ELTYLPTNASLSFLQDIQEVQGVVLAHNOVQVPLQRLRVGTQTFEDNYALAVLDNG 120

Qy 121 DPLNTPPTVGTASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHNKOLA 180

Db 121 DPLNTPPTVGTASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHNKOLA 180

Qy 181 LTLIDNRSRACHPCSPMKSGRCWGESSEDCOSLRTVCAGCARGCKGPIPTDCCHEQC 240

Db 181 LTLIDNRSRACHPCSPMKSGRCWGESSEDCOSLRTVCAGCARGCKGPIPTDCCHEQC 240

Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300

Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy 301 YNYLSTDVGSCTLVCPPLHNOEVTAEQGTQRCCKSPCARVCYGLGMQYIKANSKFTGIT 360

Db 301 YNYLSTDVGSCTLVCPPLHNOEVTAEQGTQRCCKSPCARVCYGLGMQYIKANSKFTGIT 360

Qy 361 ELEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPEQLQVFEETLEETGYLYISAWPDSL 420

Db 361 IQEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPEQLQVFEETLEETGYLYISAWPDSL 420

Qy 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLIGISWGLRLSRLSGLALIHNTLHLCFVHTV 480

Db 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLIGISWGLRLSRLSGLALIHNTLHLCFVHTV 480

Qy 481 PWQLFRNPQALLHTANRPEDCEVGEGLACHQLCARGHCHWGPQTQVCNCSQFLRQEC 540

Db 481 PWQLFRNPQALLHTANRPEDCEVGEGLACHQLCARGHCHWGPQTQVCNCSQFLRQEC 540

Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEPAQCACAHYKDPFCVCARC 600

Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEPAQCACAHYKDPFCVCARC 600

Qy 601 PSGVKPDLSTVMPWKPPDEBACQPCPINCTHSCVDLDDKGPAPQASPLTSTVSAVVG 660

Db 601 PSGVKPDLSTVMPWKPPDEBACQPCPINCTHSCVDLDDKGPAPQASPLTSTVSAVVG 660

Qy 661 ILVWVLGVVFGILIKERQKIRKYTMRLRLQSTELVEPLTPSGAMPNQAMRILKETEL 720

Db 661 ILVWVLGVVFGILIKERQKIRKYTMRLRLQSTELVEPLTPSGAMPNQAMRILKETEL 720

Qy 721 RKVKVLGSGAFGQYIKA-----NSKFGITELVLRNTSPKANKEILDEAYVMAGVGS 773

Db 721 RKVKVLGSGAFGQYIKGWIPTDGENVK-IPVAIKVLRNTSPKANKEILDEAYVMAGVGS 779

QY 774 PYVSRLLIGICLTSTVQLVQTMPLPYGCLLDHVRENRGRGLSGDILLNWCQIAKMSYLEDV 833
 Db 780 PYVSRLLIGICLTSTVQLVQTMPLPYGCLLDHVRENRGRGLSGDILLNWCQIAKMSYLEDV 839
 QY 834 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMALESILRRRF 893
 Db 840 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMALESILRRRF 899
 QY 894 THQSDVMSYGVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVVMIMVKCW 953
 Db 900 THQSDVMSYGVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVVMIMVKCW 959
 QY 954 MIDSECPRERELVSERMRARDQRFVWVQNEEDLGPASPLDSTFYRSLLDEDDMGDLVD 1013
 Db 960 MIDSECPRERELVSERMRARDQRFVWVQNEEDLGPASPLDSTFYRSLLDEDDMGDLVD 1019
 QY 1014 AEEVLVPQOQFFCPDPAPGAGMVHRRSSSTRSGGGDLTLGLEPSEERAPRSLAPSE 1073
 Db 1020 AEEVLVPQOQFFCPDPAPGAGMVHRRSSSTRSGGGDLTLGLEPSEERAPRSLAPSE 1079
 QY 1074 GAGSDVFDGLGMAAKGLOSLPTHDPSPLOQYSEDPTVPLPSETDGYVAPLTCSPQPEY 1133
 Db 1080 GAGSDVFDGLGMAAKGLOSLPTHDPSPLOQYSEDPTVPLPSETDGYVAPLTCSPQPEY 1139
 QY 1134 VNQPDVVRPQPSREGPLPAARPAAGATLERAKTLLSPKNGVGVKDVFAFGGAVENPEYLT 1193
 Db 1140 VNQPDVVRPQPSREGPLPAARPAAGATLERAKTLLSPKNGVGVKDVFAFGGAVENPEYLT 1199
 QY 1194 QGGAAPQHPHPPAFSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLGLDVVP 1249
 Db 1200 QGGAAPQHPHPPAFSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 14
 AAU77114
 ID AAU77114 standard; Protein; 1255 AA.
 XX AC AAU77114;
 XX DT 05-JUN-2002 (first entry)
 XX DE Human Her-2/neu polypeptide.
 XX KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;
 KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
 KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
 KW Hodgkin's lymphoma; T cell therapy.
 XX OS Homo sapiens.
 XX PN WO200213847-A2.
 XX PD 21-FEB-2002.
 XX PF 13-AUG-2001; 2001WO-US25408.
 XX PR 14-AUG-2000; 2000US-0638280.
 XX PR 28-SEP-2000; 2000US-0675904.
 XX PA (CORI-) CORIXA CORP.
 XX PI Gaiger A, Cheever MA, Hand-zimmermann S;
 XX WI; 2002-280741/32.
 XX DR N-PSDB; ABK10730.
 XX PT Inhibiting haematological malignancy development by administering
 PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
 PT encoding the polypeptide, or antigen presenting cells expressing the
 PT polypeptide.
 XX PS Disclosure; Page 71-74; 74pp; English.
 XX

CC The invention relates to a method for inhibiting development of
 CC haematological malignancy in a patient by administering a polypeptide
 CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
 CC encoding the polypeptide. Antigen presenting cells that express the
 CC protein can also be administered. The sequences are used for inhibiting
 CC development of haematological malignancy such as acute myelogenous
 CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
 CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
 CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
 XX SQ Sequence 1255 AA;
 Query Match 97.4%; Score 6602; DB 23; Length 1255;
 Best Local Similarity 97.2%; Pred. No. 0;
 Matches 1221; Conservative 6; Mismatches 21; Indels 8; Gaps 2;
 QY 1 METAAALCRWGLLALLPPGAASQVCTGTDMLKRLPASPTHLDMLRHLYOGQVVOGNL 60
 Db 1 METAAALCRWGLLALLPPGAASQVCTGTDMLKRLPASPTHLDMLRHLYOGQVVOGNL 60
 QY 61 ELYTLPTNASLSFLQDIQEVQGVLIHAHQVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
 Db 61 ELYTLPTNASLSFLQDIQEVQGVLIHAHQVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
 QY 121 DPLNNTPTVPGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIHFNKNOLA 180
 Db 121 DPLNNTPTVPGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIHFNKNOLA 180
 QY 181 LTLIDNTRSPACHPCSPMCKSGSCWGESSEDCOSLTRVCAGGCARCKGLPTDCCHEQC 240
 Db 181 LTLIDNTRSPACHPCSPMCKSGSCWGESSEDCOSLTRVCAGGCARCKGLPTDCCHEQC 240
 QY 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 QY 301 YNYLSTDVGSCTLVCPPLHNQEVTAEDGTQCEKSKPCARVCYGLGQHYIKANSKFTGIT 360
 Db 301 YNYLSTDVGSCTLVCPPLHNQEVTAEDGTQCEKSKPCARVCYGLGQHYIKANSKFTGIT 360
 QY 361 ELEFAGCKIFGSLAFIPESFDGDPASNTAPLOPEQLQVPELLEEITGYLIYISAWPDSL 420
 Db 361 IQEFAGCKIFGSLAFIPESFDGDPASNTAPLOPEQLQVPELLEEITGYLIYISAWPDSL 420
 QY 421 DLSVFQNLQVIRGRIILHNGAYSLTLQGLIGLSWGLSLRELGSGLALIHNTLHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRIILHNGAYSLTLQGLIGLSWGLSLRELGSGLALIHNTLHLCFVHTV 480
 QY 481 PWDQLFRNPQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQCVNCSQFLRQEC 540
 Db 481 PWDQLFRNPQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQCVNCSQFLRQEC 540
 QY 541 VEECRVLOGLPREYVNAHCLPCHPCQONGSVTCFGEADOCVACAHYKDPFVCVARC 600
 Db 541 VEECRVLOGLPREYVNAHCLPCHPCQONGSVTCFGEADOCVACAHYKDPFVCVARC 600
 QY 601 PSGVKPDLSPYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCFAEORASPLTISVAVVG 660
 Db 601 PSGVKPDLSPYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCFAEORASPLTISVAVVG 660
 QY 661 ILVVVLGVVVFGLIKRRQKIRKYTMRLQLQETELVEPLTPSGAMPNQAKMRLKETEL 720
 Db 661 ILVVVLGVVVFGLIKRRQKIRKYTMRLQLQETELVEPLTPSGAMPNQAKMRLKETEL 720
 QY 721 RKVKVLGSGAFGQYIYA-----NSKFTGITELVIRENTSPKANKEILDEAVTMAGVGS 773
 Db 721 RKVKVLGSGAFGQYIYKGIWIPDGENYK-IPVAIKVIRENTSPKANKEILDEAVTMAGVGS 779
 QY 774 PYVSRLLIGICLTSTVQLVQTMPLPYGCLLDHVRENRGRGLSGDILLNWCQIAKMSYLEDV 833
 Db 780 PYVSRLLIGICLTSTVQLVQTMPLPYGCLLDHVRENRGRGLSGDILLNWCQIAKMSYLEDV 839
 QY 834 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMALESILRRRF 893

||||| 840 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRF 899
Qy THQSDVMSYGVVWELMTFGAKPDGIPAREIPDLLEKGERLPPOPICTIDVYMWKCV 953
Db THQSDVMSYGVVWELMTFGAKPDGIPAREIPDLLEKGERLPPOPICTIDVYMWKCV 959
Qy MIDSECRPRELVSFSESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGLVD 1013
Db MIDSECRPRELVSFSESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGLVD 1019
Qy ABEYLVPQGFCCPDPAFGAGMWHHRSSSTRSGGDLTLGLEPSEERAPRSLAPSE 1073
Db ABEYLVPQGFCCPDPAFGAGMWHHRSSSTRSGGDLTLGLEPSEERAPRSLAPSE 1079
Qy GAGSDVFDGLGMAAGLQSLPHDPSPLQRYSEDPTVLPSTDGYVAPLTCSPQPEY 1133
Db GAGSDVFDGLGMAAGLQSLPHDPSPLQRYSEDPTVLPSTDGYVAPLTCSPQPEY 1139
Qy VNQPDVPRQPPSPREGPLPAARPAATLERAKTSLPGKNGVVDVFAFGGAVENPEYLT 1193
Db VNQPDVPRQPPSPREGPLPAARPAATLERAKTSLPGKNGVVDVFAFGGAVENPEYLT 1199
Qy QGGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYGLDVPV 1249
Db QGGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYGLDVPV 1255

RESULT 15

AAR39568
ID AAR39568 standard; Protein; 1433 AA.
XX AC AAR39568;
XX DT 07-FEB-1994 (first entry)
XX DE Sequence of c-erbB-2 tumour antigen.
XX KW Tumour antigen; c-erbB-2; glycoprotein.
XX OS Homo sapiens.
XX PN WO9316185-A.
XX PD 19-AUG-1993.
XX PF 05-FEB-1993; 93MO-US01055.
XX PR 06-FEB-1992; 92US-0831967.
XX PA (CETU) CETUS ONCOLOGY CORP.
XX PA (CREA-) CREATIVE BIOMOLECULES INC.
XX PI Houston LL, Huston JS, Oppermann H, Ring DB;
XX WPI; 1993-272889/34.
XX DR N-PSDB; AAQ46083.
XX CC New single chain Fv polypeptide binding to C-erbB-2 tumour
XX PT antigen - for imaging or treating breast or ovarian cancer etc.
XX PS Disclosure; pages 48-54; 87pp; English.
XX CC c-erbB-2 refers to a protein antigen expressed on the surface of
XX CC tumour cells. Such as breast and ovarian tumour cells, which is an
XX CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
XX CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents
XX CC the location of a stop codon in AAQ46083.
XX SQ Sequence 1433 AA;
Query Match 96.8%; Score 6559; DB 14; Length 1433;
Best Local Similarity 96.7%; Pred. No. 0;

Matches 1214; Conservative 8; Mismatches 26; Indels 8; Gaps 2;
Qy 1 METAAALCRWGLLALLPPGAASTQVCTGDMKRLRLPASPTHLDMLRHLHQGVQVVOGNL 60
Db 1 METAAALCRWGLLALLPPGAASTQVCTGDMKRLRLPASPTHLDMLRHLHQGVQVVOGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQGVYLAHNOVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQGVYLAHNOVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Qy 121 DPLNTPVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILWKDIIPHKNQOLA 180
Db 121 DPLNTPVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILWKDIIPHKNQOLA 180
Qy 181 LTLIDNRSRACHPCSPMKCGSESSDCSLTRTVCAAGCARCKGKPLPTDCCHEQC 240
Db 181 LTLIDNRSRACHPCSPMKCGSESSDCSLTRTVCAAGCARCKGKPLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVSGCTLVCPHNOEVTABDGTQRCCKSPCARVCYGLGMOYIKANSKFTGIT 360
Db 301 YNYLSTDVSGCTLVCPHNOEVTABDGTQRCCKSPCARVCYGLGMOYIKANSKFTGIT 360
Qy 361 ELFEAGCKTIFGSLAFPLPESFDGDPASNTAPLOEOLQVPELLEETIGYLIYISAWPDSL 420
Db 361 IQBFAGCKRIFGSLAFPLPESFDGDPASNTAPLOEHLQVPELLEETIGYLIYISAWPDSL 420
Qy 421 DLSVFQNLQVIRGRIILHNGAYSLTLOGLGTSWGLSLRSLRGLSGLALIHNTLHCFVHTV 480
Db 421 DLSVFQNLQVIRGRIILHNGAYSLTLOGLGTSWGLSLRSLRGLSGLALIHNTLHCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCHWGPGTQCVNCSQFLRQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCHWGPGTQCVNCSQFLRQEC 540
Qy 541 VEECRVLOGLPREYVNAHCLPCHPCOPONGSVTCFGEADOCVACAHYKDPFPCVARC 600
Db 541 VEECRVLOGLPREYVNAHCLPCHPCOPONGSVTCFGEADOCVACAHYKDPFPCVARC 600
Qy 601 PSGVKPDLSPYMPIWKPDEEGACQPCINCHSCVDLDDKGCAPAEORASPLTISAVVG 660
Db 601 PSGVKPDLSPYMPIWKPDEEGACQPCINCHSCVDLDDKGCAPAEORASPLTISAVVG 660
Qy 661 ILLVVLGVVFGILIKRQOKIRKYTWRLLOETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVVLGVVFGILIKRQOKIRKYTWRLLOETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLGSGAFQGYIKA-----NSKFIGITELVRENTSPKANKEILDEAYVMAGVGS 773
Db 721 RKVKVLGSGAFQGYIKGIWIPDGENVK-IPVAIKVIRENTSPKANKEILDEAYVMAGVGS 779
Qy 774 PYVSRLIGICLTSTVOLVTOLMPYGLLDHVRNRRGLSQDLLNWCMIKAGMSYLEDV 833
Db 780 PYVSRLIGICLTSTVOLVTOLMPYGLLDHVRNRRGLSQDLLNWCMIKAGMSYLEDV 839
Qy 834 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRF 893
Db 840 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRF 899
Qy 894 THQSDVMSYGVVWELMTFGAKPDGIPAREIPDLLEKGERLPPOPICTIDVYMWKCV 953
Db 900 THQSDVMSYGVVWELMTFGAKPDGIPAREIPDLLEKGERLPPOPICTIDVYMWKCV 959
Qy 954 MIDSECRPRELVSFSESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGLVD 1013
Db 960 MIDSECRPRELVSFSESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGLVD 1019
Qy 1014 ABEYLVPQGFCCPDPAFGAGMWHHRSSSTRSGGDLTLGLEPSEERAPRSLAPSE 1073
Db 1020 ABEYLVPQGFCCPDPAFGAGMWHHRSSSTRSGGDLTLGLEPSEERAPRSLAPSE 1079

Qy	1074	GAGSDVFDGLGMAAKGLQSLPTHDPSPLOQYSEDPTVPLPSETDGYVAPLTCSPQPEY	1133
Db	1080	GAGSDVFDGLGMAAKGLQSLPTHDPSPLOQYSEDPTVPLPSETDGYVAPLTCSPQPEY	1139
Qy	1134	VNQPDVVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTTP	1193
Db	1140	VNQPDVVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTTP	1199
Qy	1194	QGGAAPQHPHPPAFSPAFDNLIYYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVPV	1249
Db	1200	QGGAAPQHPHPPAFSPAFDNLIYYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVPV	1255

Search completed: July 22, 2003, 09:16:55
Job time : 42.7573 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 22.9062 Seconds
(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-5-25-14
Perfect score: 6812
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6694	98.3	1255	1 A24571	protein-tyrosine k
2	5879.5	86.3	1254	2 I48161	p-185 precursor -
3	5876	86.3	1260	1 TVRTNU	protein-tyrosine k
4	3132	46.0	1210	1 GQHUE	epidermal growth f
5	3111	45.7	1210	2 A53183	epidermal growth f
6	3086.5	45.3	1223	1 TVCHLV	epidermal growth f
7	2968.5	43.6	1308	2 A47253	epidermal growth f
8	2689	39.5	1166	1 S06142	protein-tyrosine k
9	2408	35.3	1342	2 A36223	kinase-related tra
10	2327.5	34.2	1339	2 JC4387	epidermal growth f
11	1786.5	25.9	698	1 TVFVLV	protein-tyrosine k
12	1703	25.0	604	1 TVYUHV	protein-tyrosine k
13	1652.5	24.3	1330	1 GQFFE	epidermal growth f
14	1647	24.2	544	2 S35745	protein-tyrosine k
15	1640	24.1	545	2 S00727	kinase-related tra
16	1623	23.8	540	2 B4776	epidermal growth f
17	1621	23.8	540	1 TVFVEB	protein-tyrosine k
18	1505	22.1	644	2 A36325	epidermal growth f
19	1297	19.0	1323	2 E88257	protein let-23 [im
20	1297	19.0	1374	2 S70712	protein-tyrosine k
21	1189	17.5	1369	2 S70713	protein-tyrosine k
22	1175	17.2	1717	1 A45558	epidermal growth f
23	1118	16.4	527	2 A42032	epidermal growth f
24	953.5	14.0	843	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	protein-tyrosine k
26	754.5	11.1	311	2 S13808	protein-tyrosine k
27	724	10.6	1363	2 T43220	insulin-like growt
28	702	10.3	1382	1 INHUR	insulin receptor p
29	699.5	10.3	1607	2 T43212	insulin-like growt

ALIGNMENTS

RESULT 1

A24571
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
N;Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein erb
C;Species: Homo sapiens (man)
C;Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999
C;Accession: A24571; A25491; A44188; B44188; I59509; I57622
R;Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;
Nature 319, 230-234, 1986
A;Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth
A;Reference number: A24571; MUID:86118663; PMID:3003577
A;Accession: A24571
A;Molecule type: mRNA
A;Residues: 1-1255 <YAM>
A;Cross-references: GB:X03363; NID:g31197; PIDN:CAA27060.1; PID:g31198
R;Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A;Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epider
A;Reference number: A25491; MUID:86016729; PMID:2995967
A;Accession: A25491
A;Molecule type: DNA
A;Residues: 737-1031 <SHM>
A;Cross-references: GB:M11767; NID:gi82163; PIDN:AAA35808.1; PID:g553282
R;Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg,
Science 230, 1132-1139, 1985
A;Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromo
A;Reference number: A44188; MUID:86070181; PMID:2999974
A;Accession: A44188
A;Molecule type: DNA
A;Residues: 740-910 <COU1>
A;Cross-references: GB:M12036; NID:gi83988; PIDN:AAA35978.1; PID:gi83989
A;Accession: B44188
A;Molecule type: mRNA
A;Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>
A;Cross-references: GB:M11730; NID:gi83986
R;King, C.R.; Kraus, M.E.; Aaronson, S.A.
Science 229, 974-976, 1985
A;Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
A;Reference number: I59509; MUID:85272597; PMID:2992089
A;Accession: I59509
A;Molecule type: DNA
A;Status: translated from GB/EMBL/DBJ
A;Residues: 832-909 <REX>
A;Cross-references: GB:I29395; NID:g459807; PIDN:AAA35809.1; PID:g459808
R;Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987
A;Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptiona
A;Reference number: I57622; MUID:87286898; PMID:3039351
A;Accession: I57622
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-191 <TAL>

insulin receptor p
insulin receptor p
protein-tyrosine k
insulin receptor -
insulin receptor-r
insulin receptor-r
insulin-like growt
insulin-like growt
insulin receptor -
insulin receptor -
protein-tyrosine k
protein-tyrosine k
tyrosine kinase Mp
mouse developmenta
protein-tyrosine k
protein-tyrosine k

A;Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C;Genetics:
A;Gene: GDB:ERBB2; NGL; NEU; HER-2
A;Cross-references: GDB:120613; OMIM:164870
A;Map position: 17q21.1-17q21.1
A;Introns: 25/1; 75/3; 147/1; 883/3
A;Note: the list of introns is incomplete
C;Function:
A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
inase
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F;22-653/Domain: extracellular #status predicted <EXT>
F;70-304/Domain: EGF receptor extracellular domain repeat <EE1>
F;395-605/Domain: EGF receptor extracellular domain repeat <EE2>
F;654-675/Domain: transmembrane #status predicted <TM>
F;676-1255/Domain: intracellular #status predicted <INT>
F;718-983/Domain: protein kinase homology <KIN>
F;726-734/Region: protein kinase ATP-binding motif
F;68,124,187,259,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;753/Active site: Lys #status predicted
F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 98.3%; Score 6694; DB 1; Length 1255;
Best Local Similarity 98.3%; Pred. No. 5,46-266;
Matches 1237; Conservative 5; Mismatches 9; Indels 8; Gaps 2;

Qy 1 MELAALCRWGLLLALLPGGAASCTVCFNNFTVSFVLRVVKVSASHLE----LYQGQGVV 56
Db 1 MELAALCRWGLLLALLPGGAASCTVCTGTDMK----LRLPASPETHLMDRLHLYQGQGVV 56

Qy 57 QGNLELYLPTNASLSFLQDIQEVQGVYLIHNNQVQVPLQRLRIVRGTLQFEDNYALAV 116
Db 57 QGNLELYLPTNASLSFLQDIQEVQGVYLIHNNQVQVPLQRLRIVRGTLQFEDNYALAV 116

Qy 117 LDNGDPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKIDIFHN 176
Db 117 LDNGDPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKIDIFHN 176

Qy 177 NQALTATIDNRSRACHPCSPCKSGRCWGSESSDCOSLTRVTCAGGCARCKGLPTDCC 236
Db 177 NQALTATIDNRSRACHPCSPCKSGRCWGSESSDCOSLTRVTCAGGCARCKGLPTDCC 236

Qy 237 HEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
Db 237 HEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296

Qy 297 TACPYNVLTSDVGSCITVCPHNOQVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVAV 356
Db 297 TACPYNVLTSDVGSCITVCPHNOQVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVAV 356

Qy 357 TSANIQSFAGCKTIFGSLAFIPESFDGDPASNTAPLOEQVPEETIEITGYLYISAMP 416
Db 357 TSANIQSFAGCKTIFGSLAFIPESFDGDPASNTAPLOEQVPEETIEITGYLYISAMP 416

Qy 417 DSLPDLVSFQNLQVIRGRILHNGAYSLLTQGLGTSWLGRLSRLGSLALIHNTLHCLF 476
Db 417 DSLPDLVSFQNLQVIRGRILHNGAYSLLTQGLGTSWLGRLSRLGSLALIHNTLHCLF 476

Qy 477 VHTVPVQDLFRNPHOALLHTANREDEVCVGEGLACHQLCARGHCWGPGPTQCVNCSQFLR 536
Db 477 VHTVPVQDLFRNPHOALLHTANREDEVCVGEGLACHQLCARGHCWGPGPTQCVNCSQFLR 536

Qy 537 GOECVEECRVLOGLPREVYNARHCLPCHPECPQNGSVTCFPGPADQCVACHYKDPFPC 596
Db 537 GOECVEECRVLOGLPREVYNARHCLPCHPECPQNGSVTCFPGPADQCVACHYKDPFPC 596

Qy 597 VARCPGSKVPDLSYMPNWKFPDEBEGACQPCPINCTHSCVDLDDKGCQPAEQASPLTSIVS 656
Db 1 VARCPGSKVPDLSYMPNWKFPDEBEGACQPCPINCTHSCVDLDDKGCQPAEQASPLTSIVS 656

Db 597 VARCPGSKVPDLSYMPNWKFPDEBEGACQPCPINCTHSCVDLDDKGCQPAEQASPLTSIVS 656
Qy 657 AVVGIILVVVLGVVFGILIKRQOKIRKYTMRLQLQETELVEPLTPSGAMPNQAQRILK 716
Db 657 AVVGIILVVVLGVVFGILIKRQOKIRKYTMRLQLQETELVEPLTPSGAMPNQAQRILK 716
Qy 717 ETELKRVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRRENTSPKANKEILDEAYVMAG 776
Db 717 ETELKRVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRRENTSPKANKEILDEAYVMAG 776
Qy 777 VGSPPYVRLGICLSTVOLVTLQMPYGCGLLDHVRNRRGLSGQDLLNMCQIAKMSYL 836
Db 777 VGSPPYVRLGICLSTVOLVTLQMPYGCGLLDHVRNRRGLSGQDLLNMCQIAKMSYL 836
Qy 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMALSILR 896
Db 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMALSILR 896
Qy 897 RRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYIMV 956
Db 897 RRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYIMV 956
Qy 957 KCMIDSECRPRELVSFERNARDPQRFVVIQNEDELGPASPLDSTFYRSLLDEDDMGD 1016
Db 957 KCMIDSECRPRELVSFERNARDPQRFVVIQNEDELGPASPLDSTFYRSLLDEDDMGD 1016
Qy 1017 LVDABEYLVPQGFPCPDPAAGAGVHRRHSSSTRSGGDLTLGLEPSEERAPRPLA 1076
Db 1017 LVDABEYLVPQGFPCPDPAAGAGVHRRHSSSTRSGGDLTLGLEPSEERAPRPLA 1076
Qy 1077 PSEGAGSDVFDGLGMAAKGLOSLTHDPSPLQRYSEDTVPPLPSETDGYVAPLTCSPQ 1136
Db 1077 PSEGAGSDVFDGLGMAAKGLOSLTHDPSPLQRYSEDTVPPLPSETDGYVAPLTCSPQ 1136
Qy 1137 PEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLPQKNGVQKDVFAFGAVENPEY 1196
Db 1137 PEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLPQKNGVQKDVFAFGAVENPEY 1196
Qy 1197 LTPQGGAAQPPHPPPAFSPAFDNLVYWDQDPPRGAAPPSTFKGTPTAENPEYLGLOVPV 1255
Db 1197 LTPQGGAAQPPHPPPAFSPAFDNLVYWDQDPPRGAAPPSTFKGTPTAENPEYLGLOVPV 1255

RESULT 2
I48161
p-185 precursor - golden hamster
C;Species: Mesocricetus auratus (golden hamster)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C;Accession: I48161
R;Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa
Gene 140, 251-255, 1994
A;Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A;Reference number: I48161; MID:94193007; PMID:7908275
A;Accession: I48161
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1254 <RES>
A;Cross-references: GB:D16295; NID:G493236; PIDN:BA03801.1; PID:G747595
C;Genetics:
A;Gene: neu
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP
F;718-983/Domain: protein kinase homology <KIN>
F;726-734/Region: protein kinase ATP-binding motif

Query Match 86.3%; Score 5879.5; DB 2; Length 1254;
Best Local Similarity 86.2%; Pred. No. 8,9e-233;
Matches 1085; Conservative 60; Mismatches 105; Indels 9; Gaps 3;

Qy 1 MELAALCRWGLLLALLPGGAASCTVCFNNFTVSFVLRVVKVSASHLE----LYQGQGVV 56
Db 1 MELAALCRWGLLLALLPGGAASCTVCTGTDMK----LRLPASPETHLMDRLHLYQGQGVV 56

```
QY 57 QGNLELYLPNLSLFLQDIQEVQGVVLIHNOVRQVPLQRLRIVRGTLQFEDNYALAV 116
DB 57 QGNLELYLPANATLSFLQDIQEVQGYMLIAHSQVRHVPLQRLRIVRGTLQFEDKYALAV 116
QY 117 LQNGDPLNNTTPTVGTASPGGLRELQLRLSLTEILKAGGVLIQHNPOLCYQDTILWKDIFHN 176
DB 117 LQNGDPLNNTTPTVGTASPGGLRELQLRLSLTEILKAGGVLIQHNPOLCYQDTILWKDIFHN 176
QY 117 LQNGDPLNNTTPTVGTASPGGLRELQLRLSLTEILKAGGVLIQHNPOLCYQDTILWKDIFHN 176
DB 117 LQNGDPLNNTTPTVGTASPGGLRELQLRLSLTEILKAGGVLIQHNPOLCYQDTILWKDIFHN 176
QY 177 NQLALTLIDITNRSRACHPCSPMCKGRCWGSESDCSLRTVTCAGGCARCKGPLPTDCC 236
DB 177 NQLALTLIDITNRSRACHPCSPMCKGRCWGSESDCSLRTVTCAGGCARCKGPLPTDCC 236
QY 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMNPPEGRYTFGASCV 296
DB 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMNPPEGRYTFGASCV 296
QY 297 TACPNYLSLTDVGSCTLVCPLNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRV 356
DB 297 TACPNYLSLTDVGSCTLVCPLNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRV 356
QY 357 TSANIQEFAGCKKIFGSLAPLPEFDGDPASNTAPLOQLQVPELLEEITGYLIYISWP 416
DB 357 TSANIQEFAGCKKIFGSLAPLPEFDGDPASNTAPLOQLQVPELLEEITGYLIYISWP 416
QY 417 DSLPDLVSFQNLQVIRGILHNGAYSITLQIGLISWGLSLRLSRLGSLALIHNTLCLF 476
DB 417 DSLPDLVSFQNLQVIRGILHNGAYSITLQIGLISWGLSLRLSRLGSLALIHNTLCLF 476
QY 477 VHTVPDQLPRNPHQALLHTANRPEDECVCBGLACHOLCARGHCWGPGTQVCNCSOFLR 536
DB 477 VHTVPDQLPRNPHQALLHTANRPEDECVCBGLACHOLCARGHCWGPGTQVCNCSOFLR 536
QY 537 GQECVECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQCVACAHYKDPFPC 596
DB 537 GQECVECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQCVACAHYKDPFPC 596
QY 597 VARCPGVKDLSTWPTWKPDDEGACQPCPINTHSCVDLDDKGCAPAEQASPLTISVS 656
DB 597 VARCPGVKDLSTWPTWKPDDEGACQPCPINTHSCVDLDDKGCAPAEQASPLTISVS 656
QY 657 AVGILLVVLGVVFGILIKRQOKIRKYMRLLQETELVEPLTPSGAMPNQAMRLK 716
DB 657 AVGILLVVLGVVFGILIKRQOKIRKYMRLLQETELVEPLTPSGAMPNQAMRLK 716
QY 717 ETELKRVKVLGSGAFGYWYGIWIPDGENVKIPVAIKVLRNTSPKANKELIDAYVMAG 776
DB 717 ETELKRVKVLGSGAFGYWYGIWIPDGENVKIPVAIKVLRNTSPKANKELIDAYVMAG 776
QY 777 VGSYPVSRLLGICLTSTVQLVTLQMPYGCLLDHVRENRLGSGODLLNWCMIAGMSYL 836
DB 777 VGSYPVSRLLGICLTSTVQLVTLQMPYGCLLDHVRENRLGSGODLLNWCMIAGMSYL 836
QY 837 EDVRLVHRDLAARNVLKSNHVKITDGLARLLIDIDETEHADGGKVPKWMALLESILR 896
DB 837 EDVRLVHRDLAARNVLKSNHVKITDGLARLLIDIDETEHADGGKVPKWMALLESILR 896
QY 897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMV 956
DB 897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMV 956
QY 957 KCMWIDSECRPRFRELVSFSESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDMDGD 1016
DB 957 KCMWIDSECRPRFRELVSFSESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDMDGD 1016
QY 1017 LVDAEEYLVQOQGFCDPAPGAGMVHRRHSSTSSGSGDLTLGLPESEEEAPRSPILA 1076
DB 1017 LVDAEEYLVQOQGFCDPAPGAGMVHRRHSSTSSGSGDLTLGLPESEEEAPRSPILA 1076
QY 1077 PSEGAGSDVFDGDLGMAAGLQSLPTHDPSPLOQYSEDPTVPLPSETDGVVAPLTCSPQ 1136
DB 1077 PSEGAGSDVFDGDLGMAAGLQSLPTHDPSPLOQYSEDPTVPLPSETDGVVAPLTCSPQ 1136
QY 1137 PEYVNPQDVRPQPPSPREGPLPAARAPAGATLERAKTLSPGKNGVYVKOVFAFGGAVENPEY 1196
```

```
DB 1137 PEYVNPQDVRPQPPSPREGPLPAARAPAGATLERAKTLSPGKNGVYVKOVFAFGGAVENPEY 1196
QY 1197 LTPQGGAAPOPPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
DB 1197 LTPQGGAAPOPPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1254
QY 1197 LTPQGGAAPOPPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1254
DB 1197 LTPQGGAAPOPPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1254
RESULT 3
TVRTNU
protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
C:Accession: A24562; A61204
R:Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1996
A:Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.
A:Reference number: A24562; MUID:86118662; PMID:3945311
A:Accession: A24562
A:Molecule type: mRNA
A:Residues: 1-1260 <BAR>
A:Cross-references: EMBL:X03362; NID:G56745; PIDN:CAA27059.1; PID:G56746
R:Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen,
Carcinogenesis 12, 1975-1978, 1991
A:Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no
2-thiazolyl]formamide or N-methyl-N-nitrosourea.
A:Reference number: A61204; MUID:92035293; PMID:1682063
A:Accession: A61204
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 637-663, 'V', 665-702 <NAS>
A:Note: authors translated the codon GCA for residue 25 as Val
C:Genetics:
A:Gene: neu
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosph
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>
F:658-680/Domain: transmembrane #status predicted <TMN>
F:723-988/Domain: protein kinase homology <KIN>
F:731-739/Region: protein kinase ATP-binding motif
F:711,263,535,576,634/Binding site: carboxydrate (Asn) (covalent) #status predicted
F:691/Binding site: phosphate (Thr) (covalent) #status predicted
F:758/Active site: Lys #status predicted
F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted
Query Match 86.3%; Score 5876; DB 1; Length 1260;
Best Local Similarity 86.2%; Pred. No. 1.2e-232;
Matches 1087; Conservative 54; Mismatches 110; Indels 10; Gaps 4;
QY 1 MELAALCRWGLLLALLPPGAASSTQVCTFNNFTVSFWLVPKVSASHLE---LYQGQGVV 56
DB 4 MELAANCRWGLLLALLPPGIAGTQVCTGDMK---LRLPASPETHLDMRLHLYQGQGVV 59
QY 57 QGNLELYLFTNASLSFLQDIQEVQGVVLIHNOVRQVPLQRLRIVRGTLQFEDNYALAV 116
DB 60 QGNLELYTFANASLSFLQDIQEVQGVVLIHNOVRQVPLQRLRIVRGTLQFEDKYALAV 119
QY 117 LQNGDPLNNTTPTVGTASPGGLRELQLRLSLTEILKAGGVLIQHNPOLCYQDTILWKDIFHK 175
DB 120 LQNRDPQDNVAASPTGRTPEGLRELQRLSLTEILKAGGVLIQHNPOLCYQDMVLMKQVFRK 179
QY 176 NQALTLIDITNRSRACHPCSPMCKGRCWGSESDCSLRTVTCAGGCARCKGPLPTDC 235
DB 180 NQALPVDIDITNRSRACHPCSPMCKGRCWGSESDCSLRTVTCAGGCARCKGRLPTDC 239
QY 236 CHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMNPPEGRYTFGASC 295
DB 240 CHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMNPPEGRYTFGASC 299
QY 296 VTACPNYLSLTDVGSCTLVCPLNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRV 355
DB 300 VTTCPNYLSLTDVGSCTLVCPNNQEVTAEDGTQRCCKSPCARVCYGLGMEHLRGARA 359
```


A;Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C;Genetics: :
A;Gene: GDB:ERBB2; NGL; NEU; HER-2
A;Map position: 17q21.1-17q21.1
A;Introns: 25/1; 75/3; 147/1; 883/3
A;Note: the list of introns is incomplete
C;Function: :
A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
inase
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F;22-653/Domain: extracellular #status predicted <EXT>
F;70-304/Domain: EGF receptor extracellular domain repeat <EE1>
F;395-605/Domain: EGF receptor extracellular domain repeat <EE2>
F;654-675/Domain: transmembrane #status predicted <TM>
F;676-1255/Domain: intracellular #status predicted <INT>
F;718-983/Domain: protein kinase homology <KIN>
F;726-734/Region: protein kinase ATP-binding motif
F;68.124,187,259,530,571,629/Binding site: carboxylate (Asn) (covalent) #status predicted
F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;753/Active site: Lys #status predicted
F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 97.4%; Score 6619; DB 1; Length 1255;
Best Local Similarity 96.9%; Pred. No. 1.8e-265;
Matches 1221; Conservative 6; Mismatches 17; Indels 16; Gaps 2;

Qy 1 MELAAALCRWGLLLALLPPGAASCTCTGDMKRLPASPETHDMLRLHYQGCQVQGNL 60
Db 1 MELAAALCRWGLLLALLPPGAASCTCTGDMKRLPASPETHDMLRLHYQGCQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIEVOGYVLIHAHQVQVQLRIRVGRQTLFEDNYALAVLNG 120
Db 61 ELTYLPTNASLSFLQDIEVOGYVLIHAHQVQVQLRIRVGRQTLFEDNYALAVLNG 120
Qy 121 DPLNTTPTVTGASPGGLRELQRLSLEILKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA 180
Db 121 DPLNTTPTVTGASPGGLRELQRLSLEILKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA 180
Qy 181 LTLIDTNRSRACHPCSPMCKSRGWESSEDCQSITRTVCAGGCARCKGPIPTDCCHQC 240
Db 181 LTLIDTNRSRACHPCSPMCKSRGWESSEDCQSITRTVCAGGCARCKGPIPTDCCHQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNLTSTDVGSCTLVCLPHNQEVTAEDGTQRCCKSPCARVCYGLGMQYIKANSKFIGIT 360
Db 301 YNLTSTDVGSCTLVCLPHNQEVTAEDGTQRCCKSPCARVCYGLGMHELRVAVTSAN 360
Qy 361 ELEFAGCKKIFGSLAFLPESPDGASNTAPLOPSOLOVFTLEBITGYLIYSAMPDSLP 420
Db 361 IQEFAGCKKIFGSLAFLPESPDGASNTAPLOPSOLOVFTLEBITGYLIYSAMPDSLP 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLSLTQLGLISWGLRLSRELGSGLALIHNNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLSLTQLGLISWGLRLSRELGSGLALIHNNTHLCFVHTV 480
Qy 481 PWDQLFRPHQALLTANRPEDECYVGEGLACHOLCARGHCWGPQTQCVCNSQFIRGQEC 540
Db 481 PWDQLFRPHQALLTANRPEDECYVGEGLACHOLCARGHCWGPQTQCVCNSQFIRGQEC 540
Qy 541 VEECRVQLGLPREYNARHCLPCHPECOPNGSVTCFGEADQCACAHYKDPPEVCARC 600
Db 541 VEECRVQLGLPREYNARHCLPCHPECOPNGSVTCFGEADQCACAHYKDPPEVCARC 600
Qy 601 PSKVDPDLSYMPINWPKFDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTISI 660
Db 601 PSKVDPDLSYMPINWPKFDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTISI 660

Db 601 PSKVDPDLSYMPINWPKFDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTISI -----I 655
Qy 661 SKFTGI-----TELKROOKIRKYTMRELLOETELVEPLTPSGAMPNQAOMRIL 709
Db 656 SAVVGILLVVVLGVVFGILIKRROOKIRKYTMRELLOETELVEPLTPSGAMPNQAOMRIL 715
Qy 710 KETELRKVKVLGSGAFGTGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMA 769
Db 716 KETELRKVKVLGSGAFGTGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMA 775
Qy 770 GVSGPYVRLGLICLTSTVOLVTLQMPYGLLHVHNRGRGLSQDILLNMCWQAKGMSY 829
Db 776 GVSGPYVRLGLICLTSTVOLVTLQMPYGLLHVHNRGRGLSQDILLNMCWQAKGMSY 835
Qy 830 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWALELIL 889
Db 836 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWALELIL 895
Qy 890 RRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMIM 949
Db 896 RRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMIM 955
Qy 950 VKCMWIDSECRPRELVSEFSEMRWDPQRFVVIQNEEDLGPASPLDSTFYSRLLEDDEDMG 1009
Db 956 VKCMWIDSECRPRELVSEFSEMRWDPQRFVVIQNEEDLGPASPLDSTFYSRLLEDDEDMG 1015
Qy 1010 DLVDAEYLYPQOQFFCPDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEEAPRSP 1069
Db 1016 DLVDAEYLYPQOQFFCPDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEEAPRSP 1075
Qy 1070 APSEGAGSDVFDGDLGMAAKGLQSLTHDPSPQLQRYSEDPVPLPSETDGYVAPLTCSP 1129
Db 1076 APSEGAGSDVFDGDLGMAAKGLQSLTHDPSPQLQRYSEDPVPLPSETDGYVAPLTCSP 1135
Qy 1130 QPEYVNPQDVHPQPPSPREGPLPAARPAAGATLERAKTILSPKNGVWVDVFAFGAVENPE 1189
Db 1136 QPEYVNPQDVHPQPPSPREGPLPAARPAAGATLERAKTILSPKNGVWVDVFAFGAVENPE 1195
Qy 1190 YLTPQGGAAQPHPPPPAFSPAFDNLVYDODPPERGAPPSTFKGTPTAENPEYLGLDVVP 1249
Db 1196 YLTPQGGAAQPHPPPPAFSPAFDNLVYDODPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 2
TVRTNU
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
C;Accession: A24562; A61204
R;Bartmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A;Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.
A;Reference number: A24562; MUID:86118662; PMID:3945311
A;Accession: A24562
A;Molecule type: mRNA
A;Residues: 1-1260 <BAR>
A;Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746
R;Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen,
Carcinogenesis 12, 1975-1978, 1991
A;Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no m
2-thiazolyl]formamide or N-methyl-N-nitrosourea.
A;Reference number: A61204; MUID:92035293; PMID:1682063
A;Accession: A61204
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 637-663,'V',665-702 <MAS>
A;Note: authors translated the codon GCA for residue 25 as Val
C;Genetics: :
A;Gene: neu
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:10:23 ; Search time 22.179 Seconds
(without alignments)
5413.772 Million cell updates/sec

Title: SEQ4-632-652-12
Perfect score: 6796
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1249
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6619	97.4	1255	1 A24571	protein-tyrosine k
2	5860	86.2	1260	1 TVRTWU	protein-tyrosine k
3	5833.5	85.8	1254	2 I48161	p-185 precursor
4	3115	45.8	1210	1 GQHUE	epidermal growth f
5	3082	45.4	1210	2 A53183	epidermal growth f
6	3051.5	44.9	1223	1 TVCHLV	epidermal growth f
7	2925.5	43.0	1308	2 A47253	epidermal growth f
8	2635	38.8	1166	1 S06142	protein-tyrosine k
9	2398.5	35.3	1342	2 A36223	kinase-related tra
10	2297.5	33.8	1339	2 JC4387	epidermal growth f
11	1723.5	25.4	698	1 TVFVLH	protein-tyrosine k
12	1660	24.4	604	1 TVYVUH	protein-tyrosine k
13	1617.5	23.8	1330	1 GQFE	epidermal growth f
14	1604	23.6	544	2 S35745	protein-tyrosine k
15	1597	23.5	545	2 S00727	kinase-related tra
16	1580	23.2	540	2 B47776	protein-tyrosine k
17	1578	23.2	540	1 TVFVFB	protein-tyrosine k
18	1509	22.2	644	2 A36325	epidermal growth f
19	1282	18.9	1323	2 E88257	epidermal growth f
20	1282	18.9	1374	2 S70712	protein-tyrosine k
21	1190	17.5	1369	2 S70713	protein-tyrosine k
22	1148	16.9	1717	1 A45558	epidermal growth f
23	1126	16.6	527	2 A42032	epidermal growth f
24	975.5	14.4	843	2 A27131	epidermal growth f
25	806.5	11.9	346	2 S13807	protein-tyrosine k
26	754.5	11.1	311	2 S13808	protein-tyrosine k
27	711	10.5	1363	2 T43220	insulin-like growt
28	681	10.0	1382	1 INHUR	insulin receptor p
29	679	10.0	1607	2 T43212	insulin-like growt

RESULT 1

A24571

protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
N;Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein erbB
C;Species: Homo sapiens (man)
C;Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999
C;Accession: A24571; A25491; A44188; B44188; I59509; I57622
R;Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;
Nature 319, 230-234, 1986
A;Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth f.
A;Reference number: A24571; MUID:86118663; PMID:3003577
A;Accession: A24571
A;Molecule type: mRNA

A;Residues: 1-1255 <YAM>

A;Cross-references: GB:X03363; NID:g31197; PIDN:CAA27060.1; PID:g31198

R;Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.

Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A;Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epiderm

A;Reference number: A25491; MUID:86016729; PMID:2995967

A;Accession: A25491

A;Molecule type: DNA

A;Residues: 737-1031 <SEM>

A;Cross-references: GB:M11767; NID:g182163; PIDN:AAA35808.1; PID:g553282

R;Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg, P.

Science 230, 1132-1139, 1985

A;Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromos

A;Reference number: A44188; MUID:86070181; PMID:2999974

A;Accession: A44188

A;Molecule type: DNA

A;Residues: 740-910 <COU1>

A;Cross-references: GB:M12036; NID:g183988; PIDN:AAA35978.1; PID:g183989

A;Accession: B44188

A;Molecule type: mRNA

A;Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>

A;Cross-references: GB:M11730; NID:g183986

R;King, C.R.; Kraus, M.H.; Aaronson, S.A.

Science 229, 974-976, 1985

A;Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A;Reference number: I59509; MUID:85272597; PMID:2992089

A;Accession: I59509

A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 832-909 <REX>

A;Cross-references: GB:I29395; NID:g459807; PIDN:AAA35809.1; PID:g459808

R;Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987

A;Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptional

A;Reference number: I57622; MUID:87286898; PMID:3039351

A;Accession: I57622

A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-191 <TAL>

```
QY 638 DDKCPAEQASPLTSIVSAVV-GILLVVVLGVVFGILIKEROOKIRKYTWRRLLQETEL 696
Db 58 GLEGCP---NGSKTFSIAAGVVGGLLVVVGGLGIGLYLRRR-HIVRRTLURLLQEREL 113
QY 697 VEPLTPSGAMPNQAMRILKETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLR 756
Db 114 VEPLTPSGEAPNQAHLRLILKETEPKVKVVLGFGAGFTYKGLWIPEGEKVTIPVAIKELR 173
QY 757 ENTSPKANKEILDEAYVMAGVGSPIVSRLLGICLTSTVQLTQMLPYCCLLDHVRENRR 816
Db 174 EATSPKANKEILDEAYVMASVNDPHVRCLLGICLTSTVQLITQMLPYCCLLDYIREHKDN 233
QY 817 LGSODLLNWCQIAGMSYLEDVRLVHRDLAARNVLKSPNHVKITDPGLARLLDIDETE 876
Db 234 IGSQYLLNWCQIAGMNYLEERHLVHRDLAARNVLVKTPODVKITDFGLAKQLGADEKE 293
QY 877 YHADGGKVPICKMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLE 936
Db 294 YHAEKGKVPICKMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIIPASEISSVLE 353
QY 937 KGERLPQPPICTIDVYIMVWKWMDSECRPRFRELSEFSEMRARDPORFVVIQ-NEDLG 995
Db 354 KGERLPQPPICTIDVYIMVWKWMSDADSRPKFRELIAEFSKMARDPPRYLVIQDERMH 413
QY 996 PASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFPCPDAPGAGGMVHHRHRSSTRSG 1055
Db 414 LPSPTDSKFYRTLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
QY 1056 GGDLTGLLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMAKGLQLPHTDPSPLQ 1110
Db 455 -----SRTPLLSLSLTSNNSATNCIDRNGG-----H----- 481
QY 1111 RYSEDPTVPLPSETDGYVAPLTCSPQPEYVQPDVVRPQPPSPREGPLPAARPAGAT-LER 1169
Db 482 -----PVREDGFL-----PAPEYVNO--LMPKKPSTAMVQNOIYNYISLTAISK 523
QY 1170 AKTSPGKNGVVKDVFAGGAVENPEYL 1197
Db 524 LPMDSRYQN-----SHSTAVDNPEYL 544
```

Search completed: July 22, 2003, 09:10:12
Job time : 31.9062 secs

```

Db 709 MRHVNQYATGPGY-----CAASPPRSKITTANLD-----VNMFIITGAV 749
Qy 663 LVVVLGVVFGI-LIKRROOKIRKYT--NRRLLQETELVEPLTPSGAMPNOQMRLKETE 719
Db 750 LVPTTICLVVYIICRQKAKKEIVKTMALSGREDSEPLRPSNIGANLCKLRIVKDAE 809
Qy 720 LRKVKVLSGAFGTGKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVGS 779
Db 810 LRKGVLMGAFGRVYKGVWVPEGENVKIPVAIKELLKSTGAESSEEFLEAYIMASEEH 869
Qy 780 PYVSELLGICLTSTVOLATMPLPYCCLLDHVRNRRGRIGSQDLLNWCQIAGKMSYLEDV 839
Db 870 VNLKLLAVCMSSQMLLTQMLPLCCLLDYVRNRRDKIGSKALLNWSTQIAGKMSYLEEK 929
Qy 840 RLVRHDLAARNVLVK--SPNHVKITDFGLARLLDIDETEHADGCKVPIKWMALLESILR 896
Db 930 RLVRHDLAARNVLVRLLAGEDH-----DFGLAKLLSSDSNEYKAAGRKPIKWLALCEIRN 985
Qy 897 RFRTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMV 956
Db 986 RVFTSKSDVWAFGVTIWEILLTFGQRPHENIPAKDIPDLIEVGLKLEQPEICSLDIYCTILL 1045
Qy 957 KWMIDSECRPRFRELSEFERMARDPORFVVIQNEIDLQ--PASPLDSTFVRSLLDD-- 1012
Db 1046 SCHWLDAAAMRPTFKOLTTFVAFEFARDPGRYLAILGDKFTRLPA-----YTSQDEKDLI 1098
Qy 1013 -DMGDLVDAEEVLVPOQGFPCDPAPGAGGVHHRHRSSTRSGGDLTLGLEPSEERAP 1071
Db 1099 RKLAPTTDGSIAIAKPDYDLOPKAALGPS-----HRTDCT-----DEMP 1137
Qy 1072 -----RSLAPSEAGSDVFDG---DLGMAAKGLQSLTHDPSPLOQYSEDPTVPLPS 1122
Db 1138 KLNRYCKDPNSKNSSTGDERDSSAREVGVGNLR-----LDLPV 1176
Qy 1123 ETDGVAPLTSCPOPEYVNDVPRQPPSPREGPLPAARPAAGATLERAKTILSPQKNGVVK 1182
Db 1177 DEDDYLMP-TCQPGNNNNNN-----NPNQNNAAVGAAGYM----- 1214
Qy 1183 DVFAFGGAVENPEYL---TPQGAAPQPH-----PPAFSP-AFDN 1219
Db 1215 DLIGVPSVDNPEYLLNNAQTILGVGSEPIPTQTIGIPVMGGPGTMEVKVMPGSEPTSSDH 1274
Qy 1220 LYYWD 1224
Db 1275 EYND 1279

```

RESULT 14

S35745
 protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
 C:Species: avian erythroblastosis virus
 C:Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
 C:Accession: S35745
 R:Vennstrom, B.
 submitted to the EMBL Data Library, March 1993
 A:Reference number: S35743
 A:Accession: S35745
 A:Molecule type: DNA
 A:Residues: 1-544 <VEN>
 A:Cross-references: EMBL:X12707
 C:Genetics:
 A:Gene: erbB
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; oncogene, phosphotransferase; transforming protein; tyrosine-specific P
 F:135-400/Domain: protein kinase homology <KIN>
 F:143-151/Region: protein kinase ATP-binding motif
 F:170/Active site: Lys #status predicted

Query Match 23.8%; Score 1617; DB 2; Length 544;
 Best Local Similarity 54.3%; Pred. No. 1.5e-59;
 Matches 341; Conservative 70; Mismatches 125; Indels 92; Gaps 15;

```

Qy 578 GPEADQCACAHYKDPFPCVAVRCPOYIKANSKFIGITELPDEGACQPCPINCTHSCVDL 637
Db 1 GP--DHCWKCAHFTDGPCHVKACFAGVLGENDTL-VMKYADANAVCQLCHPNCTRGCKGP 57
Qy 638 DDKCPAEQASPLTISVSAVV-GILLVVLGVVFGIILIKRROOKIRKYIMRRLLOETEL 696
Db 58 GLEQCP---NGSKTPSTAAGVVGGLCLLVVVGIGIGLYLRRR-HIVRKRTRLRLLOEREL 113
Qy 697 VEPLTPSGAMPNOQMRLKETELRKVKVLSGAFGTGKGIWIPDGENVKIPVAIKVLR 756
Db 114 VEPLTPSGEAPNOAHILKETEFKKVKVLSGAFGTGKGIWIPDGENVKIPVAIKELR 173
Qy 757 ENTSPKANKEILDEAYVMAGVGSYVSRLLGICLTSTVOLATMPLPYCCLLDHVRNRRGR 816
Db 174 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVOLITQMLPYGCLLDYIREHKDN 233
Qy 817 LGSODLLNWCQIAGKMSYLEDLVRDLAARNVLKSPNHVKITDFGLARLLDIDETE 876
Db 234 IGSQYLLNWCQIAGKMNYLEERHMHVHRLAARNVLKTPQHVKITDFGLAKOLGADEKE 293
Qy 877 YHADGGKVPKWMALLESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLE 936
Db 294 YHAGGGKVPKWMALLESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASSEISVLE 353
Qy 937 KGERLPQPPICTIDVYIMVVKMWIDSECRPRFRELSEFERMARDPORFVVIQ-NEDLG 995
Db 354 KGERLPQPPICTIDVYIMVVKMWSDADSRPKRELIAEFKWARDPPRVILVIOGDERMH 413
Qy 996 PASPLDSTFVRSLLDDMGDLVDAEYLVPOQGFPCDPAPGAGGVHHRHRSSTRSG 1055
Db 414 LPSPTDSKFTYRLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
Qy 1056 GGDITLGLPSEERAPSL-----APSEGAGSDVFDGLGMAAKGLQSLPTHDPSPQLQ 1110
Db 455 -----SRPLLSLSATSNNSATNCIDRNG-----H----- 481
Qy 1111 RYSEDDPTVPLPSETDGYVAPLTCPOPEYVNDVPRQPPSPREGPLPAARPAAGAT-LER 1169
Db 482 -----PVREDGFL-----PAPEYVQ--LMPKFPSTAMVQNIYISLTAISK 523
Qy 1170 AKTLSPQKNGVVKDVFAFGAVENPEYL 1197
Db 524 LPIDSRYN-----SHSTAVDNPEYL 544

```

RESULT 15

S00727
 kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis virus
 C:Species: avian erythroblastosis virus
 C:Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
 C:Accession: S00727
 R:Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
 Oncogene Res. 1, 265-278, 1987
 A:Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mutant
 A:Reference number: S00727; MUID:88217326; PMID:2897102
 A:Accession: S00727
 A:Molecule type: DNA
 A:Residues: 1-545 <SCO>
 A:Cross-references: EMBL:X06943
 C:Genetics:
 A:Gene: erbB
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; phosphotransferase
 F:135-400/Domain: protein kinase homology <KIN>
 F:143-151/Region: protein kinase ATP-binding motif

Query Match 23.7%; Score 1610; DB 2; Length 545;
 Best Local Similarity 54.3%; Pred. No. 2.8e-59;
 Matches 341; Conservative 69; Mismatches 126; Indels 92; Gaps 15;

Qy 578 GPEADQCACAHYKDPFPCVAVRCPOYIKANSKFIGITELPDEGACQPCPINCTHSCVDL 637
 Db 1 GP--DHCWKCAHFTDGPCHVKACFAGVLGENDTL-VMKYADANAVCQLCHPNCTRGCKGP 57

Query Match 24.8%; Score 1688; DB 1; Length 604;
Best Local Similarity 51.7%; Pred No. 2e-62;
Matches 357; Conservative 76; Mismatches 131; Indels 126; Gaps 16;
C:Comment: This sequence is tentative because the introns have not been identified.

QY 587 CAHYKDPFCVACPCQYIKANSKFITGITEPEEGACQPCPINCTHSCVDLDDKGCAPAEQ 646
DB 3 CAHFDGPHCVKACPAVGLGENDTL-VKRYADANAVCQLCHPCNCTRGCKGLEGCP--- 58

QY 647 RASPLTSIVSAW-GILLVVLGVVFGILIKRQOKIRKYMTRRLLOTELEVEPLTPSGA 705
DB 59 NGSKTPSIAAGVGLGLLVVGLGIGLYLRR-HIVKRLRLRLQREILVEPLTPSGE 117

QY 706 MPNOAMRILKETELRKVKVLGSGAFGVYXGWIIPDCGENKIPVAIKVLRENTSPKANK 765
DB 118 APNOAHILILKETEFKKVKVLGSGAFGTYIKGLWPEGEKVKIPVAIKELRENTSPKANK 177

QY 766 EILDYAYMAGVSPYVSRLLGICLTSTVQLVTQMPYGCILLDHVRENRGLSGDILNW 825
DB 178 EILDYAYMASVDNPHVCRLLGICLTSTVQLITQMPYGCILLDIYREHKDNIGSQYLLNW 237

QY 826 CMQIAKGMVLELVDLVRDLAARNVLKSPNHVKITDPLGLARLLDIDETEVHADGKVP 885
DB 238 CVQIAKGMVLEERLVRDLAARNVLKTPHVKITDPLGLAKLGADEKETHAEGKVP 297

QY 886 IKWMALESILARRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLP 945
DB 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIAREISSVLEKGERLP 357

QY 946 ICTIDVYIMVKWMIDSECRPRFELVSEFSRMARDPQRFVIO-NEDLGASPLDSTP 1004
DB 358 ICTIDVYIMVKWMIDADSRPKFELIAEFSKMARDPRLVIOGDERMHLPSPTSKF 417

QY 1005 YRSLEDDMDGLVDAEYVLPQCFPCDPAPAGAGVHHRRHSSTSGGDDLTGLG 1064
DB 418 YRLTEEDMEDIVDAEYVLPQCFPCDPAPAGAGVHHRRHSSTSGGDDLTGLG 449

QY 1065 PSEEAAPSPL-----APSEAGSDVFDGDLGMAAGKGLQSLTHDPSPLOYSDDPTVP 1119
DB 450 -----SRTPLLSLSATSNSNSATCID-----RNCQHPVREDSFVQRYSSDPTGN 495

QY 1120 LPSET--DGYVAPLTCSPQPEVYNQFVRPQPPSPREGPLPAARPAATLERAKTLSPGK 1177
DB 496 FLEESIDDGL-----PAPEYVNO--LMPKKPSTAM----- 524

QY 1178 NGVVKDVF-----GGAVENPEYLTPOGGAAPQHPPPAPSPAFD 1218
DB 525 --VQNOIYNFISLTAKSLPMSRYQNSHSTAVDNPYL-----NTNQSPLAKTVFE 574

QY 1219 NLUYWDQPPERGAPPSTFKGTPPTAENPEY 1248
DB 575 SSPYIQSNHQ-----INLDNPDY 594

RESULT 13
GOFPE
epidermal growth factor receptor - fruit fly (Drosophila melanogaster)
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C:Species: Drosophila melanogaster
C>Date: 17-Mar-1987 #sequence revision 17-Mar-1987 #text_change 11-Jun-1999
C:Accession: A00640; A38021
R:Liivneh, E.; Glazer, L.; Segal, D.; Schlessinger, J.; Shilo, B.Z.
Cell 40, 599-607, 1985
A:Title: The Drosophila EGF receptor gene homolog: conservation of both hormone binding
A:Reference number: A00640; MUID:85124611; PMID:2982499
A:Accession: A00640
A:Molecule type: DNA
A:Residues: 1-1330<LIV>
R:Wadsworth, S.C.; Vincent III, W.S.; Bilodeau-Wentworth, D.
Nature 314, 178-180, 1985
A:Title: A Drosophila genomic sequence with homology to human epidermal growth factor re
A:Reference number: A38021; MUID:85137938; PMID:2983232
A:Accession: A38021

A:Molecule type: DNA
A:Residues: 'A',832-866,'V',868-943,'QTPSLVK'<WAD>
A:Cross-references: EMBL:X02293; NID:97922; PIDN:CAA26157.1; PID:9929565
C:Comment: This sequence is tentative because the introns have not been identified.
C:Genetics:
A:Gene: FlyBase:Egfr
A:Cross-references: FlyBase:FBgn0003731
A:Map position: 2 57F
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F:1-732/Domain: extracellular #status predicted <EXT>
F:733-764/Domain: transmembrane #status predicted <TM>
F:765-1330/Domain: intracellular #status predicted <INT>
F:808-1072/Domain: protein kinase homology <KIN>
F:816-824/Region: protein kinase ATP-binding motif
F:122,300,324,363,518,688,695,700/Binding site: carbohydrate (Asn) (covalent) #status pr
F:774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:843/Active site: Lys #status predicted
F:1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predict

Query Match 24.0%; Score 1630.5; DB 1; Length 1330;
Best Local Similarity 29.7%; Pred No. 9.8e-60;
Matches 412; Conservative 178; Mismatches 412; Indels 383; Gaps 40;

QY 80 VQGVLIHAHQVQVQLRLRIVRGTLF-----EDNYALAVLDNGDPLNNTTPTVTGASP 134
DB 38 ITNVIVIGLDIIPCTLSYRLQIIRGRTLFSLSVEEKVALFV-----TY 81

QY 135 GGLRELQRLSITEILKGVLIQRPQLCYQDTILWKDIFHKNNQALALTLIDNRSRACHP 194
DB 82 SKMYTLBIPDLRDLVNGQVGFHNHNYLCHMRTIOWSEIVNSGTDYDAYNYDTAPERECPK 141

QY 195 CSPNCKSRGVSESSDCOSLTRVCAGGCA--RCKGPLPTDCCHECAAGCTGPKHSDC 252
DB 142 CHESCTHG-CWGEFGKNCQKFKUTCSPOCAGGRCYGPKEKCHLFCAGCGCTGTQKDC 200

QY 253 LACLFHNSGICELHCPALVTYNTDTPESMPNPEGRYTFGASCVTACPYNYLSTDVGSCT 312
DB 201 IACKNFDEAVSKEECPMRKYNTTVLEINPEGKAYGATCVKECP-GHLLRDNGACV 259

QY 313 LVCPHNOEVTABDGTORCEKSCPCARVCVGLGMQVIAKANSKFIGITEL-----EPAG 366
DB 260 RSCPQDKMDKGE-----CVPCNGPCPKTC-----PGVTVLHAGNIDSPN 300

QY 367 CKKIFGSLAFUPESFDG--DPASNTA-----PQPEOLOVFETLEEITGYLYISAWPDS 418
DB 301 CTVDIGNIRILDQTFSEFDQVYANYTWGPRVPIIDPERREVSVKEITGLNTEGTPQ 360

QY 419 LPDLVSFONLQVIRGILHNGAY-SLTIQHGIGISWGLRLSRLSGGLALIHNNTHLCFV 477
DB 361 FRNLSPRNLETHGRQLMESMFAALAIKSSLSLEMRNLKQISSGSVWIQHNRDLQV 420

QY 478 HTVPWDQLFRPHQALLHTANPEDEC----- 504
DB 421 SNIRWPAIQKEPEQKVWVNNENLRADLCGKFLTILISVOHNIIMHIFAICREKWNHLLGSV 480

QY 505 ----- 504
DB 481 ORGRLLSWHGSPVYLQELQFQWHLHRRLLYLIQVINSITODKSNEHQLTDACYSPSVPT 540

QY 505 -----VG 506
DB 541 SLTITERARYIQSAGLAMELEQITARSASMRHSKTLPAEGRQVPRWFLGVCASARAGIA 600

QY 507 EGLA-----CHQICARGHCWGPGTQCVNCQFRLGQECVEECRVLQGLPREVV---N 556
DB 601 EPLAGRAVCRFCHPLCELCTNYGHEQVCSKCTHYKREQCETEC-----PADHYTDBE 654

QY 557 ARHCLPCHPECPQNGSVTCFGEADQCVACAHYK-----DPPF-----CVARCPQY 603
DB 655 QRECFQRHPEC---NG---CTGPGADCKSCRNFKLPANETGTPVNSTMNCSTKCPLE 708

QY 604 IK-ANSKFIGITELPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSIVSAVVGIL 662

Db 755 TDHMLAVGSLDHAHVRLLGLCPGSSQLVQYQLPLGSLLDHVKHQHRETLQFQLLNQGV 814
QY 828 QIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDDTEYHADGGKVPK 887
Db 815 QIAKMYYLEHSMVHRDLARNVLMKSPSQVQVADFGVADLLPDDQQLHSEAKTIK 874
QY 888 WWALESILRRPFTHOSDVNSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPIC 947
Db 875 WWALESIHFGKYTHOSDVNSYGVTVWELMTFGAEPYAGLRLAEIPDLLEKGERLAQPOIC 934
QY 948 TIDVYIMVWKCMIDSECRPRFELVSEFSRMARDPQRFVIVIONEDLGPASPLDSTFTYRS 1007
Db 935 TIDVYIMVWKCMIDENIRPTFKELANEFTRMARDPPRYLVTKRAS-GEFTP--PAAPPS 991
QY 1008 LLEDDMDGLVDAEYLVPOQGFCCPDPAAGAGGMVHRHRSSTRSGGDLTLGLEPSE 1067
Db 992 VLTIKEL-----QEALEPEL-----DLDLDEABE 1017
QY 1068 E-----EAPRSLAPSEG-----AGSDVFDGLGMGAAGLQ 1099
Db 1018 EGLATSLGALSPLPTGLTRPRGSSQLSPSSGYMPMNQSSILGEACLDSAVLGGREQFSR 1077
QY 1100 SLPHDPSPLORYSEDPTVPLPSETDGYV-----APL-----TC-----SPQPE-----Y 1139
Db 1078 FISLH-PIPRGR-----PASESEGHVTGSEAELEQKSVCSRSRSPRPRGDSAY 1129
QY 1140 VNQPDVRPQPPSPREP-----LPAARPAGATLERAKTLSP-GKNGV----- 1181
Db 1130 HSQRHLLTPVPLSPGLEEDGNGYVNDTHLGRASSREGTLSSVGLSSVLGTEED 1189
QY 1182 KDVFAGGAVENPEYLTPOGGAAPQPHPP 1210
Db 1190 ED-----EVEYEMNKRGRSP-PRPP 1209

RESULT 11
TVFVLV
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
N:Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C:Species: avian leukosis virus, ALV
C>Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C:Accession: B00643; A00643
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and processing
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: B00643
A:Molecule type: mRNA
A:Residues: 1-698 <N1>
A:Cross-references: GB:M10066; GB:M13881; NID:G211749; PIDN:AAA48763.1; PID:G211750
A:Note: in Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gallus
C:Comment: This protein is synthesized as a gag-env-erbB protein.
C:Genetics:
A:Gene: gag-env-erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific protein kinase
F:1-6/Product: gag protein (fragment) #status predicted <GAG>
F:7-59/Product: env protein (fragment) #status predicted <ENV>
F:60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F:194-459/Domain: protein kinase homology <KIN>
F:202-210/Region: protein kinase ATP-binding motif
F:229/Active site: Lys #status predicted

Query Match 25.5%; Score 1736.5; DB 1; Length 698;
Best Local Similarity 51.7%; Pred. No. 2.4e-64;
Matches 370; Conservative 80; Mismatches 141; Indels 125; Gaps 18;

QY 578 GPEADQCVACAHYKDPFCVACRQYIKANSKFIGITELPDEGACQPCPINCTHSCVDL 637
Db 60 GP--DHCMKCAHFDGHCWKACGAVLGENDTI-VNKYADANAVCOLCHNCNTRGCKGP 116
QY 638 DDKGCAPAEQASPLTISAVV-GILLVVLGVVFGVILIKRQOKIRKYMRLLOETEL 696

Db 117 GLEGP---NGSKTPSIAAGVVGGLLCLVAVVGLGILYLR--HIVRKRTRLRRLQEREL 172
QY 697 VEPLTPSGAMPQAOIRLKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVL 756
Db 173 VEPLTPSGEAPQAHRLKETEFKVKVLGSGAGFTVYKGLWIPGEKVKIPVAIKELR 232
QY 757 ENTSPKANKEILD EAYVMAGVSPYVSRLLGICLTSTVQLVTQMLPYGCLLDHVRNRR 816
Db 233 EATSPKANKEILD EAYVMASVDNPHVCELLGICLTSTVQLITQMLPYGCLLDYIREHKDN 292
QY 817 LGSODLLNWCQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDDTE 876
Db 293 IGSODLLNWCQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLAKLLGADKE 352
QY 877 YHADGGKVPKMWALESIILRRFTHOSDVNSYGVTVWELMTFGAKPYDGIPIAREIPDLLE 936
Db 353 YHADGGKVPKMWALESIILRRFTHOSDVNSYGVTVWELMTFGSKPYDGIPIAREIPDLLE 412
QY 937 KGERLPQPPICTIDVYIMVWKCMIDSECRPRFELVSEFSRMARDPQRFVIVIO-NEDLG 995
Db 413 KGERLPQPPICTIDVYIMVWKCMIDADSRPKFELAEFSKWARDPPRYLVIOGDERMH 472
QY 996 PASPLDSTFTYRSLEDDMDGLVDAEYLVPOQGFCCPDPAAGAGGMVHRHRSSTRSG 1055
Db 473 LPSPTDSKFYRTLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 513
QY 1056 GGDLTGLSEEBEAPRSL-----APSEGAGSDVFDGLGMGAAGLQSLPTHDPSPLO 1110
Db 514 -----SRTPLLSLSATSNNSATNCID-----RNGQGHVREDSPVQ 550
QY 1111 RYSDPTVPLPSET--DGYVAPLTCSPQPEYVNDVVRPQPPSPREGPLPAARAGATLE 1168
Db 551 RYSDPTGNTFLRESIDGFL-----PAPEYVNO--LMPKKPS----- 585
QY 1169 RAKTLPSPKNGVVKDVF-----AFGAVENPEYLTPOGGAAPQPHPPAF 1213
Db 586 -----TAMVQNOIYNNISLTAIKLPMDSRYQNSHSTAVDNPEYL-----NTNQSPLA 633
QY 1214 SPADFNLVYNDQ-----DPPE-----RGAPPSTFKGTPTAENPEYLGLDVP 1254
Db 634 KTVFESSPYIQSGNHQINLDNPDYQODFLPNETKPNGLLKYPAAENPEYLRVAAP 689

RESULT 12
TVYUHV
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)
C:Species: avian erythroblastosis virus
C>Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C:Accession: A00644; A38022
R:Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A:Title: The erbB gene of avian erythroblastosis virus is a member of the src gene family
A:Reference number: A00644; MUID:84026539; PMID:6313229
A:Accession: A00644
A:Molecule type: DNA
A:Residues: 1-604 <YAM>
A:Cross-references: GB:K01216; NID:G209676; PIDN:AAA42400.1; PID:G209678
R:Debure, B.; Henry, C.; Benaisse, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Martin, S.
Science 224, 1456-1459, 1984
A:Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type of
A:Reference number: A38022; MUID:84223957; PMID:6328658
A:Accession: A38022
A:Molecule type: DNA
A:Residues: 1-28, 'W', 30-139, 'F', 141-145, 'V', 147-152 <DEB>
A:Cross-references: GB:K02006
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific protein kinase
F:130-395/Domain: protein kinase homology <KIN>
F:138-146/Region: protein kinase ATP-binding motif
F:165/Active site: Lys #status predicted

```
Db 411 VFSNLTITGRSLYNRGFSLLIMKNLVNTSLGFRSLKEISAGRIYISANRQLCVHSLNW 470
Qy 483 DOLFNPQALLHTA-NRPEDECVEGLACHOLCARGHCWGPPTQCVNCSQFLRGQCV 541
Db 471 TKVLRGPTTEERLDIKHNRPRDCVAEGKVCDPCLCSSGGCGPGPQCCLSCRYSGGV 530
Qy 542 EBCRVLOGLPREYVNAHCLPCHPCQPONGSVTCFGEADOCVACAHYKDPFPFCVARCP 601
Db 531 THCNFLNEPREFAHEACFCSPCEQWEGTATNGSGSUTCAQCAHFRDGPCHVCSCP 590
Qy 602 QYIKANSKFIG----ITELPDEGACQPCPINCTHSCVDLDKDGCPAQRA-----SPLTS 653
Db 591 HGV-----LGAKGPIYKVPQVQNECRPCHENCCTCKGPELQDCLGQTLVLIGKTHLTM 644
Qy 654 IVSAVVGILLVVVLGVVGGILIKRQOKIR-KYTWRRLLLOETELVEPLTPSGAMNQAM 712
Db 645 ALTVIAG--LVIFVMLGGTFLYMRGRRIQNKRAMRYLERGESIEPLDPS-EKANKVLA 701
Qy 713 RILKETELRKVKVLGSGAGFTVYKGIWIPDGNVKI PVAIKVLRNTSPKANKEILDEAY 772
Db 702 RIFKETELRKVKVLGSGVFTVHKGVWIPEGESIKIPVCIKVIEDKSGRQSFQAVTDHML 761
Qy 773 VMAGVSPVYSLRLGICLTSTVQLVTLMPYGLLDHVHRENRGLSGDNLNMCQIAKG 832
Db 762 AIGSLDHAHIVRLGLCPGSSQLQVLYPLGSLLDHVQRHAGLGPQLLLNWGVQIAKG 821
Qy 833 MSYLEDVRLVHDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKWMMALE 892
Db 822 MYTLEHGMVHRNLAARNVLLKSPSQVQVADFGVADLLPPDKQLLYSEAKTPIKWMALE 881
Qy 893 SILRRFTHOSDWSVGVTVWELMTFGAKPYGIPAREIPDLEKGERLPQPICTIDVY 952
Db 882 SHIFGKYTHOSDWSVGVTVWELMTFGAEPYAGRLAEVDPDLEKGERLAQOICTIDVY 941
Qy 953 MIMVKCMIDSECRPRFELVSEFMRMARDPQRFVVIQNEIDLGA---SPLDSTFYRSL 1009
Db 942 VMVVKCMIDENIRPTEKELANEFTMRADPPRYLVIKRES-GPGIAPGPEPHGLTNK 1000
Qy 1010 EDDMGDLVDAEYLVQOGFCPPDPAFCAGGMVHRRHSSSTRSGGDLTLGLSP-SEE 1068
Db 1001 EREVELEPDLDLDAEED-----NLATTILGSALSPLPVGTLNRP 1041
Qy 1069 EAPRSLAPSEGAGSDVFDGLMGAAKGLQLPHTD-PSPLQRYSEDPTVPLP----- 1121
Db 1042 RGSQSLSPSSGY-NPMNQNLGESCQESAVSGSSERCPRVSLH-----PMRGLCLAS 1094
Qy 1122 SETDGYVA-----PLTCSQPE-----YVNPQDVRPQPPSPREGP----- 1156
Db 1095 ESSEGHVTGSEAELOEKVSMCRSRSRSPRPRGDSAYHSQRHSLLTPVTPLSPGLEBE 1154
Qy 1157 -----LPAARPAGATLERAKTILSP-GKNGVV-----KDVAFGGAIVENPEYLTPOGA 1203
Db 1155 DVNGVYMPDTHLKGTPSSREGTLSSVGLSSVLGTEEBED-----BEYEMNRRRH 1206
Qy 1204 APQHPPEPAPFAPFDNLYWD-----QDPPERGAPPSTFKGTPTAENPEYL 1249
Db 1207 SP-PHEPRPSSLEELGYEYWDVGSGLSASLGSTQSCPLHPVIMPATGATTPDEYEM 1263
```

RESULT 10
JC4387

epidermal growth factor receptor homolog precursor - rat
N/Alternate names: ErbB3 protein; HER3 protein
C/Species: Rattus norvegicus (Norway rat)
C/Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C/Accession: JC4387
R/Hellyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A/Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.
A/Reference number: JC4387; MUID:96096535; PMID:8522190
A/Accession: JC4387
A/Molecule type: mRNA
A/Residues: 1-1339 <HEL>

A/Cross-references: GB:U29339; NID:g915389; PID:g915390
A/Experimental source: liver
A/Note: The authors translated the codon AAC for residue 369 as Thr and GTT for residue 369 as Thr. This protein is a functional heregulin receptor that transduces signals to th C;Genetics:
A/Gene: ErbB3
C/Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
C/Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F:640-659/Domain: transmembrane #status predicted <TM>
F:705-970/Domain: protein kinase homology <KIN>
F:713-721/Region: protein kinase ATP-binding motif
F:939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr) (P)

Query Match 33.9%; Score 2304.5; DB 2; Length 1339;

Best Local Similarity 40.4%; Pred. No. 2.4e-82;
Matches 521; Conservative 169; Mismatches 43; Indels 167; Gaps 36;

```
Qy 3 LAALCRWGLLLALLPPGAA---STQVCTGTDMLKRLPASPEHLDMLRLHLYQGCVVQGN 59
Db 7 LQVLC----FLLSARGSEMNSQAVCPGTLNGLSVTGDADNQYQTLYKLYEKEVWGN 62
Qy 60 LELTYLPTNASLSFLQDIQYQGVYLAHNOVQVPLQRLRIVRGTLQFEDNYALAVLDN 119
Db 63 LEIVLTGHNADLSFLQWIREVTVAVLVAMNEFSLPLNLRVVRGTQVYDGKFAIFVM-- 120
Qy 120 GDPLNNTPTVGTGSPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNQL 179
Db 121 ---LNYNT---NSHALRQLKFTQLTEILSGVYIEKNDKLCMDTIDMRDVRVR--- 170
Qy 180 ALTIIDNRSRACHPCSPMCKSGCWSESSDDCSLTRTVCAGGC-ARCKPLPTDCHE 238
Db 171 GAEIVKNGNANCPCEHVKG-RWGGPDDCQILTKICAPQNGRCFPNPNQCCHD 229
Qy 239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTFESMNPBGRVTFGASCVTA 298
Db 230 ECAGSGSPQDTCFACRRFNDSGACVPRCPDEPLVYNKLTQLEPNPHTKYQYGGVCVAS 289
Qy 299 CPYNYLSTDVSCSTLCPHLNQEVTAEDGTQCEKCKPCARVCYL--GMQYIKANSKF 356
Db 290 CPHNFV-VDQTFVRCRPPDRKMEVD-KHGLKQVCPGGLCPKACEGTSGSGRYQTVDSN 347
Qy 357 IGITELEPAGCKIFGSLAFLESFDGDPASNTAPLOPEQLOVFEETEEITGYLYISAWP 416
Db 348 ID----GFVNTKILGNLDLITGLNVDPWHKI PALDPEKLVNVRTVREITGYLNIQSWP 403
Qy 417 DSLPDLVSFQNLQVIRGILHNGAYS-LTLOGLGISWLGRLSRLSGLALIHHTHLC 475
Db 404 PHMNFVSFSLTTIGRSLYNRGFSLLIMKNLVNTSLGFRSLKEISAGRVVISANQQLC 463
Qy 476 FVHTVPDQLFRNPHQALLHTA-NRPEDECVEGLEACHOLCARGHCWGPPTQCVNCSQF 534
Db 464 YHSLNTRLLRGSEERLDIKYDRPLGECIAEGKVCDPCLCSSGGCGWPAQGCUSCRNY 523
Qy 535 LRQGCVEECRVLOGLPREYVNAHCLPCHPCQPONGSVTCFGEADOCVACAHYKDP 594
Db 524 SREGVCVTHCFLOGEPRFVHEAQCFSCHPECLPMEGTSTYNGSGSDACARCAHFRDGP 583
Qy 595 FCVARCPQYIKANSKFIG----ITELPDEGACQPCPINCTHSC--VDLDDKGCFAEORA 648
Db 584 HCVNSCPHGI-----LGAKGPIYKVPDAQNECRPCHENCCTQCGNGPELQDCLGQAEVLM 637
Qy 649 SPLTSIVSAVVGILLVVVLGVVGGILIKRQOKIR-KYTWRRLLLOETELVEPLTPSGAMP 707
Db 638 SKPHLIVAVTVG--LAVILMILGGSFLYMRGRRIQNKRAMRYLERGESIEPLDPS-EKA 694
Qy 708 NQAOVRILKETELRKVKVLGSGAGFTVYKGIWIPDGNVKI PVAIKVLRNTSPKANKEI 767
Db 695 NKVLARIKETELRKVKVLGSGVFTVHKGIWIPGESIKIPVCIKVIEDKSGRQSFQAV 754
Qy 768 LDEAYVMAGVGSPPVYRSLGLICLTSTVQLVTLMPYGLLDHVHRENRGLSGDNLNMC 827
```

```
Db 123 YQK-PPSSP--DYVQLKQLOLSNLTLSLGGVGVKSHNPLLCNVETINWMDIVDKTNP 179
QY 180 ALTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVACAGC-ARCKGPLPTDCCHE 238
Db 180 TNNLIPAFERQCKQCDHCVCNWSWAPGPGHQKFTKLLCAEQCNRCRGKFPIDCCNE 239
QY 239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA 298
Db 240 HCAGGCTGPRATDCLACRDFNDGTCKDTPPKIYDIVSHQWQNDPNIKYTFGAACVKE 299
QY 299 CPYNYLSTVDGCTLVCPHNOEVTAEADGTQCEKSCPKACVVCYGLGMOYIKANSKFTG 358
Db 300 CSNIVVTE-GACVRSACSAGMLEVD-ENGRSKCPGDGVCPCVCGIGISL-SNTIAVN 356
QY 359 ITL-EEAGCKIFGSLAPESFDGDPASNTAPIQEOLOVFETLEETGLYLTSAWPD 417
Db 357 STNIRSFNCTKINGDIIILNRNPFEGDPHYKGTWDPPEHLNLTITVKEITGLVIMWPE 416
QY 418 SLPLDSVFONLQVIRGRILHNGAYS-LTLQGLGISWGLRSLRSLGSLALHNNHLCF 476
Db 417 NMTLSVFNQLEIIRGRITTFSGFSFVVQVRHLQWGLRSLRSLGSLALHNNHLCF 476
QY 477 VHTVPKDLFRPHQALLHTANRPEDECVEGLACHOLCARGHCHGPGTQCVNCSQFLR 536
Db 477 ANTIWRLFRSEDQSIYDART-----ENQTCNNECEDGCV-PGPTMVCVSLHVD 528
QY 537 GQCEVEECRVLQGLPREYVNRHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPC 596
Db 529 GRCVASCNLLGEPREAGVQRCVQHQECLVQTDLSLTCYGFPGANCKSAHFQDQGC 588
QY 597 VARCPOYIKANSKFTIGITELPDEGACQPCPINCTHSCVDLDDKCPAEQASPLTISVS 656
Db 589 IPRCPHGILGDGDTL-IWKYADKMGCQPCQCHQCTQGCSPGLSGCRGD-IVSHSLAVG 646
QY 657 AVVGLLVVVLGVVFGILIKRQKIKYTMRLLOETELVPLTPSGMPNQAMRLIK 716
Db 647 LVSGLLITIVALLIWLRRRIK-RKRTIRCLLQELVELVPLTPSGMPNQAFRLIK 705
QY 717 ETELKRVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAG 776
Db 706 ETEFKDRVLGSGAGTVYKGLWNPDGENIRIPVAIKVLRNTSPKANKEILDEAYVMAG 765
QY 777 VGSPYVSRLLGLCLTSTVOLVTLQMPYGLLDHVRNRLGSDQLLNCMQIAKMSYL 836
Db 766 VDHPHVCRLLGLCLTSAVOLVTLQMPYGLLDYVRHQERIGCQWLLNWCQIAKGMNYL 825
QY 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGCKVPKIKWMALESILR 896
Db 826 EERHLVHRDLAARNVLLKNPNHVKITDFGLSKLLTADEKEYQADGCKVPKIKWMALESILQ 885
QY 897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEGERLPQPPICTIDVYTMV 956
Db 886 WTYTHQSDVWSYGVTVWELMTFGSKPYDGIIPAKEIASVLENGERLPQPPICTIEVYML 945
QY 957 KCMWIDSECRPRFRELVSFEMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDMDGD 1016
Db 946 KCMWIDPSRPRFRELVSFEMARDPSKYLVIQ--NLPSLSDRRLFSRLSSDD--D 1000
QY 1017 LVDAEYLVPOGFCFPCPDAPGAGGVHHRSSSTRSGGDLTLGLPEPSEEAAPRSLA 1076
Db 1001 VVDAEYLLPYKRI-----NRQGS-----EPCI 1023
QY 1077 PSEGAGSDVFDGLGMAAGKGLSLPHTDPSPLQRYSEDPTV-PLPSETDGVVAPLTCSP 1135
Db 1024 PPTGH-----PVRENSITLRLNISDPTQNALEKLDGHH----- 1055
QY 1136 QPEYVNPQDVRPQP-----PSPRE-----GPLP-AARPAGATLERAKTLSPGKNGYVK 1182
Db 1056 --EYVNPQGSSETSRSLSDIYNPNVEDLTDGMPVSLSSQEAETNFSRPEYLTNQNLSL-- 1111
QY 1183 DVFAEGGAVENPEYLTPOGGAAPQHPAPPSAFDNLVYWDQDPPERGAPSTFKGTPPT 1242
Db 1112 -PLVSSGSMDDPDY---QAG-----YQAAF-----LPQTGALTGMGMFLPA 1148
```

```
QY 1243 AENPEYLG 1250
Db 1149 AENLEYLG 1156
```

RESULT 9

```
AJ36223
kinase-related transforming protein (erbB3) (EC 2.7.1.1-) precursor - human
C:Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C:Accession: AJ36223; I59164
R:Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A:Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal
A:Reference number: AJ36223; MUID:90083234; PMID:2687875
A:Accession: AJ36223
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1342 <KRA>
A:Cross-references: GB:M29366
R:Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, G.J.
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A:Title: Molecular cloning and expression of another epidermal growth factor receptor-re-
A:Reference number: I59164; MUID:90311312; PMID:2164210
A:Accession: I59164
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>
A:Cross-references: GB:M34309; NID:G183990; PIDN:AAA35979.1; PID:G306841
C:Genetics:
A:Gene: GDB:ERBB3; HER3
A:Cross-references: GDB:I119880; OMIM:190151
A:Map position: 12q13-12q13
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo
C:Keywords: ATP; phosphotransferase
F:707-972/Domain: protein kinase homology <KIN>
F:715-723/Region: protein kinase ATP-binding motif
```

```
Query Match 35.1%; Score 2388.5; DB 2; Length 1342;
Best Local Similarity 40.1%; Pred. No. 8.8e-91;
Matches 528; Conservative 195; Mismatches 452; Indels 143; Gaps 35;
```

```
QY 10 GLLLALLPPGAA--STOVCTGTDMLRLPASPETHLMLRHLRYOGCVVQGNLELTLYLPT 67
Db 11 GLLESLARGSEVGNQAVCPGTGLNGLSVTGAENQYQTLKYLCERCEVVGNGLEIVLTGH 70
QY 68 NASLSFLQDIOEVQGYVLIAHNQVRQVPLQRLRIRVGTQLFEDNYALAVLDNGDPLNNTT 127
Db 71 NADLSFLQWIREVTGYVLVANNEFTLPLNLRVVRGTQVYDGKFAIFVM---LNVNT 125
QY 128 PVTGASPGGLRELOLRSLRILKGGVLIQRNPOLCYODTILWKDIFHKNQALTLTDN 187
Db 126 ----NSSHALRQLRLTQLTEILSGGVYIERNDKLCHMDTIDWRDIVDRD---AEIVVKD 178
QY 188 RSRACHPCSPMKSGRCWGESSEDCQSLTRTVACAGC-ARCKGPLPTDCCHEQCAAGCTG 246
Db 179 NGRSCPPCHEVCKG-RWGFGESEDCQTLTKTIQPCQNGHCFGNPNQCCHECAGGCGS 237
QY 247 PKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPNYVLT 306
Db 238 PQDTCDFACRHFNDGACVPRCPQPLVYVNLKTFQLEPNPHTKYQYGGVCVASCAPHNFV-V 296
QY 307 DVGSCITLVCLPHNOEVTAEADGTQCEKSCPKARVCYGLGMOYIKANSKP--IGITEL- 363
Db 297 DQTSVCRACPPDKMEVD-KNGLKMCPECGGLCPKACEGTG-----SGSRFQTVDSNNIDG 350
QY 364 FAGCKKIFGSLAFLPESFDGDPASNTAPIQEOLOVFETLEETGLYLTSAWPDLSL 423
Db 351 FVNCTKLGNLDFLITGLNGDPWHKIPALDPEKLVNVRTVREITGYLNIQSWPHMNF 410
QY 424 VFQNLQVIRGRILHNGAYS-LTLQGLGISWGLRSLRSLGSLALHNNHLCFVHTVPW 482
```


FT /label= Cysteine_rich_domain
 FT 210..224
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 250..264
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 324..483
 FT /label= Ligand_binding_domain
 FT 325..339
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 369..383
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 465..479
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 484..623
 FT /label= Cysteine_rich_domain
 FT 579..593
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 624..654
 FT /label= Transmembrane_domain
 FT 632..652
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 653..667
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 655..1010
 FT /label= Tyrosine_kinase_domain
 FT 661..675
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 695..709
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 710..730
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 1011..1235
 FT /label= C-terminal_domain

WO200020027-A2.

13-APR-2000.

05-OCT-1999; 99WO-DK00525.

05-OCT-1998; 98DK-0001261.

20-OCT-1998; 98US-0105011.

(MEBI-) M & E BIOTECH AS.

Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;

Gautam A, Birk P, Karlsson G;

WPI; 2000-349917/30.

N-PSDB; AAA09455.

Inducing immune responses to weakly immunogenic, tumor associated peptide antigens for the treatment of breast and prostate cancer

Claim 62; Page 193-198; 220pp; English.

This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of Her2 can be used in the claimed method as an autovaccine to induce a CTL response. Subdominant CTL epitopes, antibody binding regions and cysteine residues involved in disulfide bonds are preserved in the immunogenized forms. Regions suitable for the insertion of foreign T helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic cell-associated peptide antigens (PA) such as those associated with cancers (self-proteins), e.g. human prostate specific membrane antigen (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b). The method comprises effecting simultaneous presentation by antigen producing cells (APCs) of the animals immune system of: (1) at least 1 CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1 B-cell group derived from the cell-associated PA; and (2) at least 1 first T helper cell group which is foreign to the animal. Analogues of human PSM, human Her2 and human/murine FGF8b comprising a substantial part of all known and predicted CTL and B-cell epitopes of the respective PA and including at least one foreign T helper epitope are also claimed. The method is used to treat prostate, prostate/breast or breast cancer when the PA is human PSM, FGF8b and Her2, respectively.

XX SQ Sequence 1255 AA;

Query Match 98.6%; Score 6678; DB 21; Length 1255;

Best Local Similarity 98.3%; Pred. No. 0;

Matches 1234; Conservative 4; Mismatches 11; Indels 6; Gaps 1;

QY 1 MELAALCRWGLLLALLPFGAASQTCT-----QVIKANSKFIGITELLYQCCVVGQNL 54
 DB 1 MELAALCRWGLLLALLPFGAASQTCTGTMKRLPASPETHLDMRLHLYQCCVVGQNL 60
 QY 55 ELTYLPTNASLSFLQDIOEVQGYVLIAHNOVRQVPLQRLIRVGTQLPEDNYALAVLDNG 114
 DB 61 ELTYLPTNASLSFLQDIOEVQGYVLIAHNOVRQVPLQRLIRVGTQLPEDNYALAVLDNG 120
 QY 115 DPLANNTPVTGASPGGLRELQRLSLTEILKGGVLIQRPQLCYQDTILMKDIFHKNNQLA 174
 DB 121 DPLANNTPVTGASPGGLRELQRLSLTEILKGGVLIQRPQLCYQDTILMKDIFHKNNQLA 180
 QY 175 LTLIDTNSRACHPCSPMKSGRCWGESSEDCQSLTRIVCAGGCARCKGPLTDCHEQC 234
 DB 181 LTLIDTNSRACHPCSPMKSGRCWGESSEDCQSLTRIVCAGGCARCKGPLTDCHEQC 240
 QY 235 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRVTFGASCVTACP 294
 DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRVTFGASCVTACP 300
 QY 295 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 354
 DB 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 360
 QY 355 IOEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPQLQVFTLEITGYLYISAMPDSL 414
 DB 361 IOEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPQLQVFTLEITGYLYISAMPDSL 420
 QY 415 DLSVFQNLQVIRGRILHNGAYSLTQGLGISMGLRSURELGSGLALIHNNTHLCFVHTV 474
 DB 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISMGLRSURELGSGLALIHNNTHLCFVHTV 480
 QY 475 PWDQLFRNPQALLHTANRPEDECYGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQBC 534
 DB 481 PWDQLFRNPQALLHTANRPEDECYGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQBC 540
 QY 535 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFQPEADQCACAHYKDPFPCVARC 594
 DB 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFQPEADQCACAHYKDPFPCVARC 600
 QY 595 PSGVKPDLISYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCAPASORASPLTSIVSAVVG 654
 DB 601 PSGVKPDLISYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCAPASORASPLTSIVSAVVG 660
 QY 655 ILLVVVLGVFGILIKRQOKIRKYTMERLLQETELVEPLTPSGAMPNQAQRILKETEL 714
 DB 661 ILLVVVLGVFGILIKRQOKIRKYTMERLLQETELVEPLTPSGAMPNQAQRILKETEL 720
 QY 715 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAAYVMAGVSP 774
 DB 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAAYVMAGVSP 780

QY 775 YVSRLLGICLTSTVQLTQMPYCCLLDHYVRENRGRGLGSODLLNWCMTAKGMSYLEDYR 834
Db 781 YVSRLLGICLTSTVQLTQMPYCCLLDHYVRENRGRGLGSODLLNWCMTAKGMSYLEDYR 840
QY 835 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNWALESILRRRT 894
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNWALESILRRRT 900
QY 895 HQSDVMSYGVTVWELMTFQAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKWM 954
Db 901 HQSDVMSYGVTVWELMTFQAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKWM 960
QY 955 IDSECPRELIVSESRWARDQRFVIVIONEDLGPASPLDSTFYRSLLDEDDMGDLVA 1014
Db 961 IDSECPRELIVSESRWARDQRFVIVIONEDLGPASPLDSTFYRSLLDEDDMGDLVA 1020
QY 1015 EYLVLPQQGFFCDDPAPGAGMVHHRSSSTRSGGDLTLGLEPSEERAPSPAPSE 1074
Db 1021 EYLVLPQQGFFCDDPAPGAGMVHHRSSSTRSGGDLTLGLEPSEERAPSPAPSE 1080
QY 1075 AGSDVFDGLGMAAGKLSLPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1134
Db 1081 AGSDVFDGLGMAAGKLSLPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1135 NQPDVREPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVAFAGGAVENPEYLTQ 1194
Db 1141 NQPDVREPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVAFAGGAVENPEYLTQ 1200
QY 1195 GGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLOV 1249
Db 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLOV 1255

RESULT 2
AAE12130
ID AAE12130 standard; Protein; 1255 AA.
XX
AC AAE12130;
XX
XX
DT 18-DEC-2001 (first entry)
XX
DE Human tyrosine kinase-type receptor, HER-2.
XX
KW Therapeutic compound; major histocompatibility complex; vaccine;
KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
KW antigen presenting cell; human; tyrosine kinase-type receptor.
XX
OS Homo sapiens.
XX
FH Key
FT Region
FT 774..782
FT /note= "Antigenic epitope"
XX
PN WO200168677-A2.
XX
XX
PD 20-SEP-2001.
XX
XX 16-MAR-2001; 2001WO-US40328.
XX
XX 16-MAR-2000; 2000US-0527487.
PR
XX (GENZ) GENZYME CORP.
XX
XX Nicolette CA;
PI
XX
XX WPI; 2001-616284/71.
DR N-PSDB; AAD19731.
XX
PT Novel synthetic therapeutic compound for inducing immune response and
PT for use in adoptive immunotherapy, has enhanced binding to major
PT histocompatibility molecules and enhanced immunoregulatory properties

XX
PS
XX
CC The invention relates to synthetic therapeutic compounds (antigenic
CC peptides) with enhanced binding to major histocompatibility complex
CC (MHC) molecules and enhanced immunoregulatory properties relative
CC to their natural counterparts. Compounds of the invention are useful
CC for inducing an immune response in a subject and for use in adoptive
CC immunotherapy. They are useful as components of anti-cancer vaccines
CC and to expand immune effector cells that are specific for cancers
CC characterised by expression of the breast cancer antigen, HER-2.
CC Polynucleotides that encode peptides of the invention are useful as
CC hybridisation probes and as primers for the detection of genes of gene
CC transcripts that are expressed in antigen presenting cells (APCs), to
CC confirm transduction of polynucleotides into host cells. The present
CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
CC of the invention are designed based on the HER-2 antigenic peptide
CC (774-782).
XX
SQ Sequence 1255 AA;
Query Match 98.6%; Score 6678; DB 22; Length 1255;
Best Local Similarity 98.3%; Pred No. 0;
Matches 1234; Conservative 4; Mismatches 11; Indels 6; Gaps 1;
QY 1 MELAALCRWGLLLALLPPGAASSTOVCT-----QYIKANSKFIGITELYQGCQVVG 54
Db 1 MELAALCRWGLLLALLPPGAASSTOVCTGTDMKRLPASPETHLDMLRLHLYQGCQVVG 60
QY 55 ELTYLPTNASLSFLQDIOEVQYVLI AHNQVQVPLQRLIRVGTQLPEDNVALAVLDNG 114
Db 61 ELTYLPTNASLSFLQDIOEVQYVLI AHNQVQVPLQRLIRVGTQLPEDNVALAVLDNG 120
QY 115 DPLNNTTPTVGASPGGLRELQRLSITELIKGVLQIQRNPOLCYQDTILWKDIFHKNQ 174
Db 121 DPLNNTTPTVGASPGGLRELQRLSITELIKGVLQIQRNPOLCYQDTILWKDIFHKNQ 180
QY 175 LTLIDNRSRACHPCSPMKSGSRGWESSEDCQSILTRTVACGGCARKGKPLTDCHEQC 234
Db 181 LTLIDNRSRACHPCSPMKSGSRGWESSEDCQSILTRTVACGGCARKGKPLTDCHEQC 240
QY 235 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAC 294
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAC 300
QY 295 YNLSLTDVGSCTLVCPHNOEVTADGTQCEKSKPCARVCYGLGMEHLREVRAVTSAN 354
Db 301 YNLSLTDVGSCTLVCPHNOEVTADGTQCEKSKPCARVCYGLGMEHLREVRAVTSAN 360
QY 355 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLQEPLOQVFEITGLYISAMPDLSLP 414
Db 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLQEPLOQVFEITGLYISAMPDLSLP 420
QY 415 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLRSLRELGSGLALIHNNTHLCFVHTV 474
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLRSLRELGSGLALIHNNTHLCFVHTV 480
QY 475 PWDQLFRPHQALLHTANRPEDECVGEGLAHOLCARGHCWGPGTQCNCQFIRGQEC 534
Db 481 PWDQLFRPHQALLHTANRPEDECVGEGLAHOLCARGHCWGPGTQCNCQFIRGQEC 540
QY 535 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVAC 594
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVAC 600
QY 595 PSGVKPDLSPYPIWKFPDEEGACQPCINCTHSCVDLDDKGPABQASPLTSIVSAVVG 654
Db 601 PSGVKPDLSPYPIWKFPDEEGACQPCINCTHSCVDLDDKGPABQASPLTSIVSAVVG 660
QY 655 ILLVVVLGVVGIILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 714
Db 661 ILLVVVLGVVGIILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720

QY 715 RKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 774
Db 721 RKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 780
QY 775 YVSRLLGICLTSTVOLVQTLMPYGCCLLDHVRENRHGLGSODLLNWCQIAKMSYLEDVR 834
Db 781 YVSRLLGICLTSTVOLVQTLMPYGCCLLDHVRENRHGLGSODLLNWCQIAKMSYLEDVR 840
QY 835 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEVHADGGKVPKMWALESIILRRFT 894
Db 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEVHADGGKVPKMWALESIILRRFT 900
QY 895 HQSDVMSYGVYVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPCTIDVYMWKCMW 954
Db 901 HQSDVMSYGVYVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPCTIDVYMWKCMW 960
QY 955 IDSECPRELYSESRWARDQRFVWIONEDLGPASPLDSTFYRSLLEDDDDMDGLVDA 1014
Db 961 IDSECPRELYSESRWARDQRFVWIONEDLGPASPLDSTFYRSLLEDDDDMDGLVDA 1020
QY 1015 EYLVVPOQGFCDPAPAGAGMWHRRSSSTRSGGDLTLGLEPSEEBAPRSLAPSEG 1074
Db 1021 EYLVVPOQGFCDPAPAGAGMWHRRSSSTRSGGDLTLGLEPSEEBAPRSLAPSEG 1080
QY 1075 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1134
Db 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1135 NQDVRPQPPSPREGPLPAARPAATLERAKTLSPOKNGVVKDVFAFGAVENPEYLTPO 1194
Db 1141 NQDVRPQPPSPREGPLPAARPAATLERAKTLSPOKNGVVKDVFAFGAVENPEYLTPO 1200
QY 1195 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGPTAENPEYLGLDVVP 1249
Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGPTAENPEYLGLDVVP 1255

RESULT 3
AAB60167
ID AAB60167 standard; Protein; 1255 AA.
XX AAB60167;
AC AAB60167;
XX
DT 03-APR-2001 (first entry)
XX
DE HER2 transgene plasmid construct encoded protein.
XX
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
XX antibody.
XX Homo sapiens.
OS Synthetic.
XX
FN WO200100244-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-US17229.
XX
PR 25-JUN-1999; 99US-0141316.
PR 16-MAR-2000; 2000US-0189844.
XX
PA (GETH) GENENTECH INC.
XX
PI Erickson S, Schwall R;
XX WPI; 2001-061962/07.
DR N-PSDB; AAF24297.
XX
PT Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
PT conjugating the antibody to a maytansinoid -
XX

PS Example 3; Fig 4; 92pp; English.
XX
CC The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ Sequence 1255 AA;

Query Match 98.6%; Score 6678; DB 22; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1234; Conservative 4; Mismatches 11; Indels 6; Gaps 1;

QY 1 MELAALCRWGLLLALLPFGAASQTCTCTGTDMKRLPASPETHLDMRLHLYQGCVVQGNL 54
Db 1 MELAALCRWGLLLALLPFGAASQTCTCTGTDMKRLPASPETHLDMRLHLYQGCVVQGNL 60
QY 55 ELTYLPTNASLFLQDIOEVQGYVLI AHNOVQVPLQRLRIVRGTQLPEDNVALAVLDNG 114
Db 61 ELTYLPTNASLFLQDIOEVQGYVLI AHNOVQVPLQRLRIVRGTQLPEDNVALAVLDNG 120
QY 115 DFLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 174
Db 121 DFLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
QY 175 LTLIDTNRACHPCSPCKSGRSGESSEDCQSLTRTVACGGCARCKGPLPTDCHEQC 234
Db 181 LTLIDTNRACHPCSPCKSGRSGESSEDCQSLTRTVACGGCARCKGPLPTDCHEQC 240
QY 235 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 294
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 295 YNLSLTDVGSCTLVCPHLNQEVTAEDGTQRCCKPCARVCYGLGMBHLREVRVTSAN 354
Db 301 YNLSLTDVGSCTLVCPHLNQEVTAEDGTQRCCKPCARVCYGLGMBHLREVRVTSAN 360
QY 355 IOEPAGCKKIFGSLAFIPESFDGDPASNTAPLOPEQLQVFFTEBITGYLIYSANPDSLP 414
Db 361 IOEPAGCKKIFGSLAFIPESFDGDPASNTAPLOPEQLQVFFTEBITGYLIYSANPDSLP 420
QY 415 DLSVFNQLQVIRGRILHNGAYSLTLQGIISWLGSLRSLRELGSGLALTHNTHLCFVHTV 474
Db 421 DLSVFNQLQVIRGRILHNGAYSLTLQGIISWLGSLRSLRELGSGLALTHNTHLCFVHTV 480
QY 475 PWDQLFRNPHQALLHTANRPEDECYGEGLACHQLCARGHCWGPPTQCVNCSQFLRGQEC 534
Db 481 PWDQLFRNPHQALLHTANRPEDECYGEGLACHQLCARGHCWGPPTQCVNCSQFLRGQEC 540
QY 535 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPEADQCVACAHYKDPFPCVARC 594
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPEADQCVACAHYKDPFPCVARC 600
QY 595 PSGVKPDLVSYPIMKFPDEEGACQPCINCTHSCVDLDDKGCPCAPORASPLTSIVSAVVG 654
Db 601 PSGVKPDLVSYPIMKFPDEEGACQPCINCTHSCVDLDDKGCPCAPORASPLTSIVSAVVG 660
QY 655 ILLVVVLGVWFGLIKRRQKIRKYTMRLRLQETELVEPLTPSGAMPNQAQMRILKETEL 714
Db 661 ILLVVVLGVWFGLIKRRQKIRKYTMRLRLQETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 715 RKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 774
Db 721 RKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 780
QY 775 YVSRLLGICLTSTVOLVQTLMPYGCCLLDHVRENRHGLGSODLLNWCQIAKMSYLEDVR 834
Db 781 YVSRLLGICLTSTVOLVQTLMPYGCCLLDHVRENRHGLGSODLLNWCQIAKMSYLEDVR 840
QY 835 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEVHADGGKVPKMWALESIILRRFT 894

Db 841 LVHRDLAARNLVKSPNNVKITDGLARLLDIDETEHADGGKVPKIMWMALESILRRRT 900
QY 895 HQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYIMVWKWM 954
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYIMVWKWM 960
QY 955 ISECRPRELIVSESRMARDPQRFVWIONEDLGPASPLDSTFYRSLLDDDDMDGLVDA 1014
Db 961 ISECRPRELIVSESRMARDPQRFVWIONEDLGPASPLDSTFYRSLLDDDDMDGLVDA 1020
QY 1015 BEVLVPQQGFPCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEERAPRPLAPSSG 1074
Db 1021 BEVLVPQQGFPCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEERAPRPLAPSSG 1080
QY 1075 AGSDVFDGLGMAAGKGLQSLPHTDPSPLQRYSESDTVPLPSETDGYVAPLTCSPQPEYV 1134
Db 1081 AGSDVFDGLGMAAGKGLQSLPHTDPSPLQRYSESDTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1135 NQPDVVRPQPPSPREGPLPAARPAAGATLERAKTILSPKNGVVKDVPFAGGAVENPEYLTPO 1194
Db 1141 NQPDVVRPQPPSPREGPLPAARPAAGATLERAKTILSPKNGVVKDVPFAGGAVENPEYLTPO 1200
QY 1195 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1249
Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 4

AAU74545
ID AAU74545 standard; Protein; 1255 AA.

AC AAU74545;

DT 23-APR-2002 (first entry)

XX Human HER2 (ErbB2) polypeptide.

XX Human; HER2; ErbB; epidermal growth factor receptor; receptor;
KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
KW glial disorder; astrocytal disorder; hypothalamic disorder;
KW glandular disorder; macrophagal disorder; epithelial disorder;
KW stromal disorder; blastocoealic disorder; inflammatory disorder;
KW angiogenic disorder; immunological disorder.

XX Homo sapiens.

XX US2002001587-A1.

XX 03-JAN-2002.

XX 16-MAR-2001; 2001US-0811123.

XX 16-MAR-2000; 2000US-189844P.

PR 05-OCT-2000; 2000US-238327P.

XX (ERIC/) ERICKSON S.

PA (SCHW/) SCHWALL R.

PA (SLIW/) SLIWKOWSKI M.

XX Erickson S, Schwall R, Sliwkowski M;

PI WPI; 2002-163686/21.

DR N-PSDB; ABK14058.

XX Treating tumour characterised by overexpression of epidermal growth
PT factor receptor, ErbB or cancer in mammal, comprises administering
PT anti-ErbB antibody-maytansinoid conjugate to the mammal -

XX Example 3; Fig 7; 93pp; English.

XX The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor
CC receptor (ErbB) and does not respond or responds poorly, to treatment
CC with an anti-ErbB antibody, comprising administering to the mammal an
CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
CC prostate and bladder, preferably breast cancer. The breast cancer is a
CC metastatic breast cancer or an aggressive form of metastatic breast
CC cancer which overexpresses ErbB2. The method is also useful for treating
CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
CC epithelial, stromal, blastocoealic, inflammatory, angiogenic and
CC immunological disorders. This sequence represents the human HER2 (ErbB2)
CC polypeptide of the invention.

XX Sequence 1255 AA;

QY Query Match 98.6%; Score 6678; DB 23; Length 1255;

Best Local Similarity 98.3%; Pred. No. 0;

Matches 1234; Conservative 4; Mismatches 11; Indels 6; Gaps 1;

QY 1 MELAALCRWELLLALLPPGAASSTQVCT-----QYIKANSKFIGITELLYQSCQVQGNL 54

Db 1 MELAALCRWELLLALLPPGAASSTQVCTGDMKRLPASPETHLDMRLHLYQSCQVQGNL 60

QY 55 ELTYLPTNASLSFLQDIOEVQGYVLIHNVQRPQRLRIVRGTQOLFEDNTALAVLDNG 114

Db 61 ELTYLPTNASLSFLQDIOEVQGYVLIHNVQRPQRLRIVRGTQOLFEDNTALAVLDNG 120

QY 115 DPLNNTTPTVTCGASPGGLRELQRLSITEILKGGVLTORNPOLCYODTILWKDIFHKNNOLA 174

Db 121 DPLNNTTPTVTCGASPGGLRELQRLSITEILKGGVLTORNPOLCYODTILWKDIFHKNNOLA 180

QY 175 LTLIDTNRSRACHPCSPMKCSRCWGESSEDCQSILTRTVTCAGGCARCKGKPLTDCCHEQC 234

Db 181 LTLIDTNRSRACHPCSPMKCSRCWGESSEDCQSILTRTVTCAGGCARCKGKPLTDCCHEQC 240

QY 235 AAGCTGPKXSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 294

Db 241 AAGCTGPKXSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300

QY 295 NYLSTDVSGCTLVCPHNOEVTAEADGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 354

Db 301 NYLSTDVSGCTLVCPHNOEVTAEADGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360

QY 355 IOEFAGCKKI FGSLAFLPESFDGDPASNTAPLOPEQLOVFTELEITGYLISAMPDLSLP 414

Db 361 IOEFAGCKKI FGSLAFLPESFDGDPASNTAPLOPEQLOVFTELEITGYLISAMPDLSLP 420

QY 415 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLRSLRELGSGLALIHNNTHLCFVHTV 474

Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLRSLRELGSGLALIHNNTHLCFVHTV 480

QY 475 PWDQLFRNPHQALLHTANRPEDECYVGBGLACHOLCARGHCWGPPTQCVNCSQFLRGQBC 534

Db 481 PWDQLFRNPHQALLHTANRPEDECYVGBGLACHOLCARGHCWGPPTQCVNCSQFLRGQBC 540

QY 535 VEECRVLOGLPREYVYVNRHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFPFCVARC 594

Db 541 VEECRVLOGLPREYVYVNRHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFPFCVARC 600

QY 595 PSGVKPDLSPYMPWKFPDDEGACQPCINCTHSCVDLDDKGCAPASQASPLTSIYSAVVG 654

Db 601 PSGVKPDLSPYMPWKFPDDEGACQPCINCTHSCVDLDDKGCAPASQASPLTSIYSAVVG 660

QY 655 ILLVVVLGVVFGILIKRRQOKIRKYMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 714

Db 661 ILLVVVLGVVFGILIKRRQOKIRKYMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

QY 715 RKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEIILDEAYVMAGVQSP 774

Db 721 RKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEIILDEAYVMAGVQSP 780

QY 775 YVSRLLIGICLTSTVOLVTQLMPYGCLLDHRVNRGRGLSGSQDILLNWCMIQAKGMSYLEDYR 834

Db 781 YVSRLLGICLTSTVQLVTQMPYGCCLLDHVRENRLGSDLLNWCQIAKMSYLEDVR 840
QY LVHRDLAARNLVKSPNHVKITDGLARLLDIDETEVHADGGKVPKKNWALESIILRRFT 894
Db 841 LVHRDLAARNLVKSPNHVKITDGLARLLDIDETEVHADGGKVPKKNWALESIILRRFT 900
QY HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMVKWM 954
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMVKWM 960
QY IDSECRPRRELVSERWARDQRFVVIQNEDLGPASPLDSTFYRSLLDDDDMDLVA 1014
Db 961 IDSECRPRRELVSERWARDQRFVVIQNEDLGPASPLDSTFYRSLLDDDDMDLVA 1020
QY 1015 EYLVFQQGFCFDPAPGAGGMVHRHRSSSTSGGGDLTLGLEPSEEAAPRSLAPSG 1074
Db 1021 EYLVFQQGFCFDPAPGAGGMVHRHRSSSTSGGGDLTLGLEPSEEAAPRSLAPSG 1080
QY 1075 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1134
Db 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1135 NQPDVPRQPPSPREGPLPAARPAAGATLERAKTSLSPGKNGVVDVFAFGAVENPEYLTPO 1194
Db 1141 NQPDVPRQPPSPREGPLPAARPAAGATLERAKTSLSPGKNGVVDVFAFGAVENPEYLTPO 1200
QY 1195 GGAAPQHPHPPAPSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLGLDVVP 1249
Db 1201 GGAAPQHPHPPAPSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 5

AAW01111
ID AAW01111 standard; Protein; 1255 AA.

XX AC AAW01111;

XX 01-JAN-1997 (first entry)

XX HER-2/neu protein.

XX HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;

XX breast cancer; ovary cancer; colon cancer; lung cancer;

XX prostate cancer; immunisation; tumour; vaccine; vector.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Domain 676..1255

XX /label= Intracellular domain

XX /note= "claimed domain, useful for immunisation"

XX W09630514-AL.

XX 03-OCT-1996.

XX 28-MAR-1996; 96WO-US01689.

XX 31-MAR-1995; 95US-0414417.

XX (UNIW) UNIV WASHINGTON.

XX Cheever MA, Disis ML;

XX WPI; 1996-455361/45.

XX N-PSDB; AAT40739.

XX DNA encoding HER-2-neu poly:peptide(s) - used for prevention or

XX treatment of malignancies with which the HER-2/neu oncogene is

XX associated

XX Claim 2; Page 56-61; 71pp; English.

XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.
XX SQ Sequence 1255 AA;
Query Match 98.5%; Score 6672; DB 17; Length 1255;
Best Local Similarity 98.2%; Pred. No. 0;
Matches 1232; Conservative 5; Mismatches 12; Indels 6; Gaps 1;
QY 1 MELAALCRWGLLLALLPPGAASTQVCT-----QYIKANSKFIGITELYQCQVVGNL 54
Db 1 MELAALCRWGLLLALLPPGAASTQVCTGTDMLRLPASPETHLMLRLHYQCQVVGNL 60
QY 55 ELTYLPTNASLSFLQDIOEVQGYVLIAHNQVQVPLQRLRIVRGTQLPEDNYALAVLDNG 114
Db 61 ELTYLPTNASLSFLQDIOEVQGYVLIAHNQVQVPLQRLRIVRGTQLPEDNYALAVLDNG 120
QY 115 DPLNNTTPTVGTASPGGLRELQRLSLEILKGGVLIQRPOLCYQDTILWKDIFHKNNOLA 174
Db 121 DPLNNTTPTVGTASPGGLRELQRLSLEILKGGVLIQRPOLCYQDTILWKDIFHKNNOLA 180
QY 175 LTLIDTNRSRACHPCSPCKSGRSGESSEDCQSLTRTVCAAGGCARCKGPLTDCCHEQC 234
Db 181 LTLIDTNRSRACHPCSPCKSGRSGESSEDCQSLTRTVCAAGGCARCKGPLTDCCHEQC 240
QY 235 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESNPNEGRTYTFGASCVTACP 294
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESNPNEGRTYTFGASCVTACP 300
QY 295 NYLSTDVSGCTLVCPHNEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRAVTSAN 354
Db 301 NYLSTDVSGCTLVCPHNEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRAVTSAN 360
QY 355 IOEFAGCKKIFGSLAFPESEFDGDPASNTAPLOPEQLQVFETLEITGYLYISAMPDSL 414
Db 361 IOEFAGCKKIFGSLAFPESEFDGDPASNTAPLOPEQLQVFETLEITGYLYISAMPDSL 420
QY 415 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGRLSRLGSLALIHNNTHLCFVHTV 474
Db 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGRLSRLGSLALIHNNTHLCFVHTV 480
QY 475 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSQFLRGQBC 534
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSQFLRGQBC 540
QY 535 VEECRVLQGLPREYVYNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFCVARC 594
Db 541 VEECRVLQGLPREYVYNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFCVARC 600
QY 595 PSGVKPDLSPYMPIWKFPDEEGACQPCPNCTHSCVDLDDKGPAPQASPLTSIVSVVG 654
Db 601 PSGVKPDLSPYMPIWKFPDEEGACQPCPNCTHSCVDLDDKGPAPQASPLTSIVSVVG 660
QY 655 ILLVVVLGVFGILLIKRQOKIRKVTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 714
Db 661 ILLVVVLGVFGILLIKRQOKIRKVTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 715 RKVKVLSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEIILDEAYVMAGVSP 774
Db 721 RKVKVLSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEIILDEAYVMAGVSP 780
QY 775 YVSRLLGICLTSTVQLVTQMPYGCCLLDHVRENRLGSDLLNWCQIAKMSYLEDVR 834
Db 781 YVSRLLGICLTSTVQLVTQMPYGCCLLDHVRENRLGSDLLNWCQIAKMSYLEDVR 840

QY 895 HQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPICTIDVYMWKCM 954
 DB 901 HQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPICTIDVYMWKCM 960
 QY 955 ISECRPRFRELVSERMRARDPQRFVIONEDLGPASPLDSTFYRSLLDDDDMDGLVDA 1014
 DB 961 ISECRPRFRELVSERMRARDPQRFVIONEDLGPASPLDSTFYRSLLDDDDMDGLVDA 1020
 QY 1015 EYLVPOQGFCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1074
 DB 1021 EYLVPOQGFCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
 QY 1075 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDTVPLPSETDGYVAPLTCSPQPEV 1134
 DB 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDTVPLPSETDGYVAPLTCSPQPEV 1140
 QY 1135 NOPDVRPQPPSPREGPLPAARPAATLERAKTLPSPKNGVVDVFAFGGAVENPEYLTPO 1194
 DB 1141 NOPDVRPQPPSPREGPLPAARPAATLERAKTLPSPKNGVVDVFAFGGAVENPEYLTPO 1200
 QY 1195 GGAAPQHPHPPAFPAFDNLYWDQPPERGAPPSTFKGTPTAENPEYLGLDVVP 1249
 DB 1201 GGAAPQHPHPPAFPAFDNLYWDQPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 7

AA221198 standard; protein; 1255 AA.

AC AA221198;

DT 12-JAN-2001 (first entry)

DE Human HER-2/neu protein.

XX Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
 KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
 KW colon cancer.

XX Homo sapiens.

XX WO200044899-A1.

XX 03-AUG-2000.

XX 28-JAN-2000; 2000WO-US02164.

XX 29-JAN-1999; 99US-0117976.

XX (CORI-) CORIXA CORP.

XX (SMIK) SMITHKLINE BEECHAM.

PI Cheever MA, Gheysen D;

XX WPI; 2000-505976/45.

DR N-PSDB; AAA89736.

XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins
 PT useful for vaccinating against breast, ovarian, colon, lung and
 PT prostate cancers -

XX Claim 52; Fig 7; 128pp; English.

XX The present sequence is the human HER-2/neu protein. It is a member of
 CC the tyrosine kinase family of receptor-like glycoproteins and shows
 CC homology to the epidermal growth factor receptor (EGFR). It probably
 CC plays a part in cell growth and/or differentiation. The HER-2/neu
 CC gene is an oncogene. An HER-2/neu fusion protein comprising a
 CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
 CC domain may be used to treat or prevent cancer by eliciting or
 CC enhancing an immune response to the HER-2/neu protein. It may be used
 CC to treat malignancies such as breast, ovarian, colon, lung and
 CC prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.
 XX
 SQ Sequence 1255 AA;
 Query Match 98.5%; Score 6672; DB 21; Length 1255;
 Best Local Similarity 98.2%; Pred. No. 0;
 Matches 1232; Conservative 5; Mismatches 12; Indels 6; Gaps 1;
 QY 1 MELAAALCRWGLLLALLPPGAASQTCTCTGTDMKLRIPASPEHLDMLRLHYOCQVQVGNL 54
 DB 1 MELAAALCRWGLLLALLPPGAASQTCTCTGTDMKLRIPASPEHLDMLRLHYOCQVQVGNL 60
 QY 55 EUTYLPNTASLSFLQDIOEVQYVLIANQVRQVPLQRLIRVGTQLPEDNVALAVLDNG 114
 DB 61 EUTYLPNTASLSFLQDIOEVQYVLIANQVRQVPLQRLIRVGTQLPEDNVALAVLDNG 120
 QY 115 DPLNNTTPTVGASPGGLRELQRLSLEILKGGVLIQRPQLCYQDTILWKDIFHKNNQLA 174
 DB 121 DPLNNTTPTVGASPGGLRELQRLSLEILKGGVLIQRPQLCYQDTILWKDIFHKNNQLA 180
 QY 175 LTLIDTNSRACHPCSPMKSGRCSGSESDCQSLTRTVAGGCARCKGPLTDCHEQC 234
 DB 181 LTLIDTNSRACHPCSPMKSGRCSGSESDCQSLTRTVAGGCARCKGPLTDCHEQC 240
 QY 235 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 294
 DB 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 QY 295 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 354
 DB 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 360
 QY 355 IOEFAGCKKIFGSLAFIPESFDGDPASNTAPLOQLOVFETLEBITGYLIISAMPDLSL 414
 DB 361 IOEFAGCKKIFGSLAFIPESFDGDPASNTAPLOQLOVFETLEBITGYLIISAMPDLSL 420
 QY 415 DLSVFQNLQVTRGRLTHNGAYSLTLQGLGISWLGSLRELGSGLALIHNNTHLCFVHTV 474
 DB 421 DLSVFQNLQVTRGRLTHNGAYSLTLQGLGISWLGSLRELGSGLALIHNNTHLCFVHTV 480
 QY 475 PWDQLFRNPQALLHTANRPEDECYGEGLACHQLCARGHCWGPGTQCVCNCSOFLRGQEC 534
 DB 481 PWDQLFRNPQALLHTANRPEDECYGEGLACHQLCARGHCWGPGTQCVCNCSOFLRGQEC 540
 QY 535 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFCVARC 594
 DB 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFCVARC 600
 QY 595 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGPABQASPLTSIVSAVVG 654
 DB 601 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGPABQASPLTSIVSAVVG 660
 QY 655 ILLVVVLGVFGILIKRRQKIRKYTMRLLLQETELVEPLTPSGAMPNQAQMRILKETEL 714
 DB 661 ILLVVVLGVFGILIKRRQKIRKYTMRLLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
 QY 715 RKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKIILDEAYVMAGVGP 774
 DB 721 RKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKIILDEAYVMAGVGP 780
 QY 775 YVSRLLGICLTSTVOLVTQMLPYGCLLDHVRNRLGSLQDLLNWCMIKAGMSYLEDVR 834
 DB 781 YVSRLLGICLTSTVOLVTQMLPYGCLLDHVRNRLGSLQDLLNWCMIKAGMSYLEDVR 840
 QY 835 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWALESIILRRFT 894
 DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWALESIILRRFT 900
 QY 895 HQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPICTIDVYMWKCM 954
 DB 901 HQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPICTIDVYMWKCM 960
 QY 955 IDSECRPRFRELVSERMRARDPQRFVIONEDLGPASPLDSTFYRSLLDDDDMDGLVDA 1014

Db 961 IDSECKPRELVSEFSRWARDPQREVVIONEDLGASPLDSTFYRSLLDDMDGLVDA 1020
QY 1015 EYLVPQOQFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1074
Db 1021 EYLVPQOQFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
QY 1075 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1134
Db 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1135 NOPDVRPQPPSREGPLPAARAGATLERAKTILSPCKNGVWVDVFAFGGAVENPEVLTPO 1194
Db 1141 NOPDVRPQPPSREGPLPAARAGATLERPKTILSPCKNGVWVDVFAFGGAVENPEVLTPO 1200
QY 1195 GGAAPOPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKTPTAENPEYLGLOVPV 1249
Db 1201 GGAAPOPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKTPTAENPEYLGLOVPV 1255

RESULT 8
AAY84780
ID AAY84780 standard; Protein; 1255 AA.
XX
AC AAY84780;
XX
DT 08-AUG-2000 (first entry)
XX
DE Amino acid sequence of the SPLICE erbb-2 receptor protein.
XX
KW SPLICE erbb-2 receptor protein; cell transformation disorder; cancer;
KW tumor cell proliferation; tissue degeneration; arthropathy;
KW bone resorption; inflammatory disease; degenerative disorder;
XX wound healing.
XX
OS Homo sapiens.
XX
PN WO2000020579-A1.
XX
PD 13-APR-2000.
XX
PF 01-OCT-1999; 99WO-CA00912.
XX
PR 02-OCT-1998; 98US-0165192.
XX
PA (UYMC-) UNIV MCMASTER.
XX
PI Muller WJ, Siegel PM;
XX
DR WPI; 2000-303768/26.
DR N-PSDB; AAA14812.
XX
PT Nucleic acid encoding an erbb 2 receptor protein designated SPLICE
PT erbb-2, inhibitors of the protein are useful for treatment of cancer -
XX
PS Claim 3; Fig 2; 60pp; English.
XX
CC The present sequence represents a SPLICE erbb-2 receptor protein. The
CC protein has an in-frame deletion of 16 amino acids, 2 of which are
CC conserved cysteine residues, compared to the unspliced protein. The
CC erbb-2 polynucleotide is used to construct probes for detecting
CC disorders of cell transformation such as cancer. Antibodies to the
CC protein may be used to detect SPLICE erbb-2 in a sample. Agents
CC (e.g. antisense oligonucleotides) which inhibit the expression of
CC SPLICE erbb-2 are useful for reducing tumor cell proliferation and
CC treating cancer. Substances which stimulate SPLICE erbb-2 are useful
CC for treating conditions involving damaged cells including conditions
CC in which degeneration of tissue occurs, such as arthropathy, bone
CC resorption, inflammatory diseases, degenerative disorders of the
CC central nervous system and wound healing.
XX
SQ Sequence 1255 AA;

Query Match 98.5%; Score 6672; DB 21; Length 1255;
Best Local Similarity 98.2%; Pred. No. 0;
Matches 1232; Conservative 5; Mismatches 12; Indels 6; Gaps 1;
QY 1 MELAALCRWGLLLALLPPGAASCTVCT-----QYIKANSKRFITGTELLYQSCQVQVGNL 54
Db 1 MELAALCRWGLLLALLPPGAASCTVCTGTDMLKRLPASPETHLDMLRHLYQSCQVQVGNL 60
QY 55 ELTYLPTNASLSFLQDIOEVGYVLIHNRQVQLRLRIVRGTQLFEDNYVALVDNG 114
Db 61 ELTYLPTNASLSFLQDIOEVGYVLIHNRQVQLRLRIVRGTQLFEDNYVALVDNG 120
QY 115 DPLNNTTPTVGPGLRELQRLSLTEILKGVLIQRPOLCYQOTILWKDIFHKQNOLA 174
Db 121 DPLNNTTPTVGPGLRELQRLSLTEILKGVLIQRPOLCYQOTILWKDIFHKQNOLA 180
QY 175 LTLIDTNRSEACHPCSPCKSGRSGESSEDCOSLTRTVCAGGCARCKGPLTDCHEOC 234
Db 181 LTLIDTNRSEACHPCSPCKSGRSGESSEDCOSLTRTVCAGGCARCKGPLTDCHEOC 240
QY 235 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 294
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 295 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVTSAN 354
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVTSAN 360
QY 355 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFEETLEBITGYLYISAMPDSUP 414
Db 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFEETLEBITGYLYISAMPDSUP 420
QY 415 DLSVFQNLQVIRGRILHNGAYSLLTQGLGISWGLRLSRLRELGSGLALIHNNTHLCFVHTV 474
Db 421 DLSVFQNLQVIRGRILHNGAYSLLTQGLGISWGLRLSRLRELGSGLALIHNNTHLCFVHTV 480
QY 475 PWDQLFRNPHQALLHTANRPEDECVEGGLACHQLCARGHCWGPGTQCVNCSQFIRGQBC 534
Db 481 PWDQLFRNPHQALLHTANRPEDECVEGGLACHQLCARGHCWGPGTQCVNCSQFIRGQBC 540
QY 535 VEECRVLOGLPREYVNRHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 594
Db 541 VEECRVLOGLPREYVNRHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600
QY 595 PSGVKPDLSPYMPWKFPEDEGACQPCINCTHSCVDLDDKGPAPQASPLTSIVSAVVG 654
Db 601 PSGVKPDLSPYMPWKFPEDEGACQPCINCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660
QY 655 ILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 714
Db 661 ILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 715 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAYVMAGVQSP 774
Db 721 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAYVMAGVQSP 780
QY 775 YVSRLLGCLSTVQVLTQMLPYGCLLDHVRNRLGSLQDLNNWCMQIAKMSYLEVDVR 834
Db 781 YVSRLLGCLSTVQVLTQMLPYGCLLDHVRNRLGSLQDLNNWCMQIAKMSYLEVDVR 840
QY 835 LVHRDLAARNVLKSPNHNKVTDFGLARLLDIDETEHADGKVPFKWMALESILRRRPT 894
Db 841 LVHRDLAARNVLKSPNHNKVTDFGLARLLDIDETEHADGKVPFKWMALESILRRRPT 900
QY 895 HQSDVMSVGVVWELMTGAKPYDGI PAREIPDLLEKGERLPPOPCTCTIDVYMIWVKWM 954
Db 901 HQSDVMSVGVVWELMTGAKPYDGI PAREIPDLLEKGERLPPOPCTCTIDVYMIWVKWM 960
QY 955 IDSECRPRFRLVSEFSRWARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDMDGLVDA 1014
Db 961 IDSECRPRFRLVSEFSRWARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDMDGLVDA 1020
QY 1015 EYLVPQOQFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1074

Db 1021 EYLVPQGFPCDPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEARPLAPSEG 1080
QY 1075 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEV 1134
Db 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEV 1140
QY 1135 NQDVRPQPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYLTQ 1194
Db 1141 NQDVRPQPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYLTQ 1200
QY 1195 GGAAPQHPPPPAFDNLVYWDQPPRGAPPSTFKGPTTAENPEYLGLDV 1249
Db 1201 GGAAPQHPPPPAFDNLVYWDQPPRGAPPSTFKGPTTAENPEYLGLDV 1255

RESULT 9
AAB85458
ID AAB85458 standard; Protein; 1255 AA.
XX AAB85458;
AC
XX
DT 25-SEP-2001 (first entry)
XX Human HER-2/neu protein.
XX Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KW oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX Homo sapiens.
XX WO200153463-A2.
XX 26-JUL-2001.
XX 19-JAN-2001; 2001WO-US01850.
XX 21-JAN-2000; 2000US-0177545.
XX (CORI-) CORIXA CORP.
XX Cheever MA, Hand-Zimmermann S;
PI WPI; 2001-476112/51.
DR N-PSDB; AA423392.
XX
XX New antigen-presenting cells, useful as vaccines for eliciting or
PT enhancing an immune response to HER-2/neu protein, particularly useful
PT for treating or preventing cancer, e.g. breast cancer -
XX
XX Claim 2; Page 41-46; 49pp; English.
XX The invention provides an isolated antigen-presenting cell, which
CC expresses at least an immunogenic portion of a polypeptide that produces
CC an immune response to HER-2/neu protein. The antigen-presenting cells are
CC useful as vaccines for eliciting or enhancing an immune response to
CC HER-2/neu protein, particularly in treating or preventing malignancies in
CC which the HER-2/neu oncogene is associated. Specifically, these are
CC useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC colon, lung or prostate cancers. The present sequence represents
CC the human HER-2/neu protein (also known as p185 or c-erbB2).
XX
SQ Sequence 1255 AA;

Query Match 98.5%; Score 6672; DB 22; Length 1255;
Best Local Similarity 98.2%; Pred. No. 0;
Matches 1232; Conservative 5; Mismatches 12; Indels 6; Gaps 1;
QY 1 MELAAALCRWGLLALLPPGAASSTQVCT-----QYIKANSKFTGITITELLYQSCVQVQNL 54
Db 1 MELAAALCRWGLLALLPPGAASSTQVCTGDMKRLRLPASPTHLDMLRHLYQSCVQVQNL 60
QY 55 ELTYLPTNASLSFLQDIQEVQVGLIAHNRQVPLQLRLIRVGTQLPEDNYALAVLDNG 114

Db 61 ELTYLPTNASLSFLQDIQEVQVGLIAHNRQVPLQLRLIRVGTQLPEDNYALAVLDNG 120
QY 115 DLANNNTVTGASPGGLRELQLRLSLEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 174
Db 121 DLANNNTVTGASPGGLRELQLRLSLEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
QY 175 LTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSLTRTVACGGCARGKPLPTDCCHEQC 234
Db 181 LTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSLTRTVACGGCARGKPLPTDCCHEQC 240
QY 235 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 294
Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 295 YNYLSTDVGSCTLVCPHNOEVTADGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 354
Db 301 YNYLSTDVGSCTLVCPHNOEVTADGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
QY 355 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOQVFETLEBITGYLIYSAMPDSL 414
Db 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOQVFETLEBITGYLIYSAMPDSL 420
QY 415 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWGLRSRLRELGLALIHNNTHLCFVHTV 474
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWGLRSRLRELGLALIHNNTHLCFVHTV 480
QY 475 PWDQLFRPHQALLHTANRDECEYGEGLACHOLCARGHCWGPGTQCWCNQSFIRGQBC 534
Db 481 PWDQLFRPHQALLHTANRDECEYGEGLACHOLCARGHCWGPGTQCWCNQSFIRGQBC 540
QY 535 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFPCVARC 594
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFPCVARC 600
QY 595 PSGVKPDLSPYMPKPFDEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIYSAVVG 654
Db 601 PSGVKPDLSPYMPKPFDEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIYSAVVG 660
QY 655 ILLVVVLGVVFGILIKRRQKIRKYMRLRLQETELVEPLTPSGAMPNQAQNRILKETEL 714
Db 661 ILLVVVLGVVFGILIKRRQKIRKYMRLRLQETELVEPLTPSGAMPNQAQNRILKETEL 720
QY 715 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVRENTSPKANKEIIDEAIVMAGVGP 774
Db 721 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVRENTSPKANKEIIDEAIVMAGVGP 780
QY 775 YVSRLLGICLTSTVOLVTQLMPYGCCLLDHVRNRLGSLQDLNWCMIAGKMSYLEDVR 834
Db 781 YVSRLLGICLTSTVOLVTQLMPYGCCLLDHVRNRLGSLQDLNWCMIAGKMSYLEDVR 840
QY 835 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMWALESILRRFT 894
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMWALESILRRFT 900
QY 895 HQSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLEKGERLPQPPICITIDVYMINVKWM 954
Db 901 HQSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLEKGERLPQPPICITIDVYMINVKWM 960
QY 955 IDSECRPRFRELVSFESRMARDPQRFVVIQNEDLGPASPDLSTFYRSLLDDDDMDGLVDA 1014
Db 961 IDSECRPRFRELVSFESRMARDPQRFVVIQNEDLGPASPDLSTFYRSLLDDDDMDGLVDA 1020
QY 1015 EYLVPQGFPCDPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEARPLAPSEG 1074
Db 1021 EYLVPQGFPCDPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEARPLAPSEG 1080
QY 1075 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEV 1134
Db 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEV 1140
QY 1135 NQDVRPQPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYLTQ 1194

Db 1141 NQPDVRQPSPREGPLPAARPAAGATLIERPKTLSPGKNGVVKDVFEGGAVENPEYLTPQ 1200

Qy 1195 GGAAAPQHPPPAFSPADNLVYWDQDPPERGAPSTFKGTPTAENPEYLGLDVVP 1249
|||||

Db 1201 GGAAAPQHPPPAFSPADNLVYWDQDPPERGAPSTFKGTPTAENPEYLGLDVVP 1255
|||||

RESULT 10

RESOLUT TO
AAG88267
ID AAG88267 standard; Protein; 1255 AA.

AC AAG88267:

DT 11-SEP-2001 (first entry)

DE HER2/neu amino acid sequence.

Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell; immune response; vaccine; tumour; cancer; cytostatic; immunostimulant; tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.

OS Homo sapiens.

XX
PN
WO200141787-A1.

14-JUN-2001

11-DEC-2000: 2000WO-US33591.

AA
PR 10-DEC-1999; 99US-0458299.

PA (EPIM-) EPIMUNE INC.

PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
PI Keogh E;

DR WPI; 2001-374995/39.

XX An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer -
XX Disclosure; Page 15; 199pp; English. PS

The present invention describes isolated prepared HER2/neu epitopes (I). Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is culture in vitro and binds to a complex of an epitope (I), bound to a human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I) and a second epitope and the peptide is less than 50 contiguous amino acids that have 100% identity with a native peptide sequence of HER2/neu; (3) a vaccine composition (III) comprising (II) and a pharmaceutical excipient; (4) an isolated nucleic acid encoding a peptide comprising (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic and immunostimulant activities, and can be used in vaccines. (I), (II) and (III) are useful for inducing cellular immune responses for the prevention and treatment of cancer. (I) and (II) are useful for monitoring or evaluating an immune response to a tumour-associated antigen when incubated with a T lymphocyte sample from a patient and detecting the presence of bound T lymphocyte to (I) or (II). Epitope based vaccines mean that immunosuppressive epitopes that may be present in whole antigens may be avoided. Selected epitopes may be combined to enhance immunogenicity. The possible pathological side effects caused by infectious agents or whole protein antigen is eliminated. The vaccine provides the ability to direct and focus an immune response to multiple selected antigens from the same pathogen. Epitope-based anti-tumour vaccines provides the opportunity to combine epitopes derived from multiple tumour-associated molecules addressing the problem of tumour-tumour variability and reducing the likelihood of tumour escape due to antigen loss. AAG88266 to AAG95121 represent amino acid sequences used in the exemplification of the present invention.

Sequence 1255 AA:

Query Match

98.5%; Score 6672; DB 22; Length 1255;

Best Local Similarity 98.2%; Pred. No.:0;		Matches 1232; Conservative		5; Mismatches		12; Indels		6; Gaps		1;	
1	MELAAALCRWGLLLALLPPGAAS	1	QV	1	QV	1	QV	1	QV	1	QV
1	MELAAALCRWGLLLALLPPGAAS	1	QV	1	QV	1	QV	1	QV	1	QV
55	ETLPTNASTSLFLODIEVOG	55	ET	55	ET	55	ET	55	ET	55	ET
61	ETLPTNASTSLFLODIEVOG	61	ET	61	ET	61	ET	61	ET	61	ET
115	DPLNNTTPVTGASPGGLRELQ	115	DPL	115	DPL	115	DPL	115	DPL	115	DPL
121	DPLNNTTPVTGASPGGLRELQ	121	DPL	121	DPL	121	DPL	121	DPL	121	DPL
175	LTLIDTNRSEACHPCSPMKSR	175	LTL	175	LTL	175	LTL	175	LTL	175	LTL
181	LTLIDTNRSEACHPCSPMKSR	181	LTL	181	LTL	181	LTL	181	LTL	181	LTL
235	AAGCTGPKHEDCLACLFHNS	235	AAG	235	AAG	235	AAG	235	AAG	235	AAG
241	AAGCTGPKHEDCLACLFHNS	241	AAG	241	AAG	241	AAG	241	AAG	241	AAG
295	YNYLSTDVGSCTLVCP	295	YNY	295	YNY	295	YNY	295	YNY	295	YNY
301	YNYLSTDVGSCTLVCP	301	YNY	301	YNY	301	YNY	301	YNY	301	YNY
355	IOEFACKKIFGSLAFLPES	355	IOE	355	IOE	355	IOE	355	IOE	355	IOE
361	IOEFACKKIFGSLAFLPES	361	IOE	361	IOE	361	IOE	361	IOE	361	IOE
415	DLSVFQNLQVIRGRILHNGA	415	DLV	415	DLV	415	DLV	415	DLV	415	DLV
421	DLSVFQNLQVIRGRILHNGA	421	DLV	421	DLV	421	DLV	421	DLV	421	DLV
475	PWDQLFRNPHQALLHTANR	475	PWD	475	PWD	475	PWD	475	PWD	475	PWD
481	PWDQLFRNPHQALLHTANR	481	PWD	481	PWD	481	PWD	481	PWD	481	PWD
535	VRECRVLOGLPREYNARHCL	535	VRE	535	VRE	535	VRE	535	VRE	535	VRE
541	VRECRVLOGLPREYNARHCL	541	VRE	541	VRE	541	VRE	541	VRE	541	VRE
595	PSGVKPDLSYPMIWKFPD	595	PSG	595	PSG	595	PSG	595	PSG	595	PSG
601	PSGVKPDLSYPMIWKFPD	601	PSG	601	PSG	601	PSG	601	PSG	601	PSG
655	ILLVVLGVVFGILIKRQOKI	655	ILL	655	ILL	655	ILL	655	ILL	655	ILL
661	ILLVVLGVVFGILIKRQOKI	661	ILL	661	ILL	661	ILL	661	ILL	661	ILL
715	RKVKVLGSGAFVYKGIWI	715	RKV	715	RKV	715	RKV	715	RKV	715	RKV
721	RKVKVLGSGAFVYKGIWI	721	RKV	721	RKV	721	RKV	721	RKV	721	RKV
775	YVSRLLIGICLSTVOLTLQ	775	YVS	775	YVS	775	YVS	775	YVS	775	YVS
781	YVSRLLIGICLSTVOLTLQ	781	YVS	781	YVS	781	YVS	781	YVS	781	YVS
835	LVHRDLAARNVLKSPNHVKI	835	LVH	835	LVH	835	LVH	835	LVH	835	LVH
841	LVHRDLAARNVLKSPNHVKI	841	LVH	841	LVH	841	LVH	841	LVH	841	LVH
895	HQSDVMSYGVTVWELMTGAK	895	HQS	895	HQS	895	HQS	895	HQS	895	HQS
901	HQSDVMSYGVTVWELMTGAK	901	HQS	901	HQS	901	HQS	901	HQS	901	HQS
955	IDSECRPRELVSSEFSRMARD	955	IDS	955	IDS	955	IDS	955	IDS	955	IDS
961	IDSECRPRELVSSEFSRMARD	961	IDS	961	IDS	961	IDS	961	IDS	961	IDS
1015	EEYLPVQGFPCDPAPGAGM	1015	EEY	1015	EEY	1015	EEY	1015	EEY	1015	

Db 1021 EYLVPOQGFCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEAAPSLAPSEG 1080
QY 1075 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPOPEYV 1134
Db 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPOPEYV 1140
QY 1135 NQPDVRQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVYKDVFAFGGAVENPEYLTPO 1194
Db 1141 NQPDVRQPPSPREGPLPAARPAAGATLERPKTLSPGKNGVYKDVFAFGGAVENPEYLTPO 1200
QY 1195 GGAAPQHPHPPAPSPAFDNLYYWDQDPPPERGAPSPSTFKGPTTAENPEYLGIDVPV 1249
Db 1201 GGAAPQHPHPPAPSPAFDNLYYWDQDPPPERGAPSPSTFKGPTTAENPEYLGIDVPV 1255

RESULT 11

AAE24067
ID AAE24067 standard; Protein; 1255 AA.
AC AAE24067;
XX
XX 23-SEP-2002 (first entry)
XX Human Her-2 protein.
DE
XX Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KW hyperproliferative disorder; prophylaxis; inflammation; antisense;
KW tumour; gene therapy; phosphorothioate backbone.
XX Homo sapiens.
OS
XX WO200222636-A1.
PN
XX 21-MAR-2002.
PD
XX 12-SEP-2001; 2001WO-US28572.
PF
XX 15-SEP-2000; 2000US-0663834.
PR
XX (ISIS-) ISIS PHARM INC.
PA
XX Bennett CF, Cowsett LM;
PI WPI; 2002-471192/50.
XX N-PSDB; AAD38904.
DR
XX Novel antisense oligonucleotide which modulates the expression of Human
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT Inflammation or to prevent infection in humans -
XX
XX Example 13; Page 95-107; 116pp; English.

CC The invention relates to antisense compounds targetted to a nucleic
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC that specifically hybridises with and inhibits the expression of Her2.
CC Antisense compounds of the invention are used for treating diseases or
CC conditions associated with Her2 such as hyperproliferative disorders
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC neural or cardiac cancer. They are also useful prophylactically e.g.
CC to prevent or delay infection, inflammation and tumour formation. The
CC invention is also used in gene therapy. The present sequence is human
CC Her-2 protein.
XX
XX Sequence 1255 AA;

Query Match 98.5%; Score 6672; DB 23; Length 1255;
Best Local Similarity 98.2%; Pred. No. 0;
Matches 1232; Conservative 5; Mismatches 12; Indels 6; Gaps 1;

QY 1 MELAALCRGGLLALLPPGAASTOVCT-----QYIKANSKFIGITELLYQCGVQGNL 54
Db 1 MELAALCRGGLLALLPPGAASTOVCTGTGDMKLRLPASPETHLMRLHYQCGVQGNL 60

QY 55 ELTYLPTNASLSFLQDIOEVQGYVLIHAHQVQVPLQRLRIVRGTQLPEDNYALAVLDNG 114
Db 61 ELTYLPTNASLSFLQDIOEVQGYVLIHAHQVQVPLQRLRIVRGTQLPEDNYALAVLDNG 120
QY 115 DPLNNTTPTVGTASPGGLRELQRLSLEILKGVLIQORNPOLCYQDTILWKDIFHKKNOLA 174
Db 121 DPLNNTTPTVGTASPGGLRELQRLSLEILKGVLIQORNPOLCYQDTILWKDIFHKKNOLA 180
QY 175 LTLIDTNSRACHPCSPCKSGRCWGESSEDCQSLTRTVACGGCARGKGPLTDCCHEOC 234
Db 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCQSLTRTVACGGCARGKGPLTDCCHEOC 240
QY 235 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 294
Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 295 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQRCBKSKPCARVCYGLGMEHLREVRVTSAN 354
Db 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQRCBKSKPCARVCYGLGMEHLREVRVTSAN 360
QY 355 IOEFAGCKKIFGSLAFLPESPDGDPASNTAPLOPELOVFETLEETITGYLISAMPDLSL 414
Db 361 IOEFAGCKKIFGSLAFLPESPDGDPASNTAPLOPELOVFETLEETITGYLISAMPDLSL 420
QY 415 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLRELGLALIHNTHLFCFVHTV 474
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLRELGLALIHNTHLFCFVHTV 480
QY 475 PWDQLFRNPQALHTANRPEDECVGEGGLACHQLCARGHCWGPQTQVCNCSOFLRGQSC 534
Db 481 PWDQLFRNPQALHTANRPEDECVGEGGLACHQLCARGHCWGPQTQVCNCSOFLRGQSC 540
QY 535 VEECEVLOGLPREYVNAHCLPCHPECOPOGNSVTCFGEADQCVACAHYKDPPECVARC 594
Db 541 VEECEVLOGLPREYVNAHCLPCHPECOPOGNSVTCFGEADQCVACAHYKDPPECVARC 600
QY 595 PSGVKPDLSPYMPKPFDEEGACQPCPINCSTHSCVDLDDKGPAPQORASPLTSISAVVG 654
Db 601 PSGVKPDLSPYMPKPFDEEGACQPCPINCSTHSCVDLDDKGPAPQORASPLTSISAVVG 660
QY 655 ILLVVVLGVFGILIKRQOKIRKYTMRLRLQETELVEPLTPSGAMPNQAQRILKETEL 714
Db 661 ILLVVVLGVFGILIKRQOKIRKYTMRLRLQETELVEPLTPSGAMPNQAQRILKETEL 720
QY 715 RKVKYLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAVVMAGVQSP 774
Db 721 RKVKYLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAVVMAGVQSP 780
QY 775 YVSRLLGICLTSTVOLVTLQMPYGCCLLDHVRNRCGLSGQDILLNWCMIKAGMSYLEDVR 834
Db 781 YVSRLLGICLTSTVOLVTLQMPYGCCLLDHVRNRCGLSGQDILLNWCMIKAGMSYLEDVR 840
QY 835 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPDKWMALESILRRRPT 894
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPDKWMALESILRRRPT 900
QY 895 HQSDVMSYGVTVWELMTGAKPYDGI PARETPDLLEKGERLPQPPICITDVMYIMVKWM 954
Db 901 HQSDVMSYGVTVWELMTGAKPYDGI PARETPDLLEKGERLPQPPICITDVMYIMVKWM 960
QY 955 IDSECRPFRELVSFMSRMARDPQRFVVIQNEIDLQSPASPLDSTFYRSLLEDDMDGLDVA 1014
Db 961 IDSECRPFRELVSFMSRMARDPQRFVVIQNEIDLQSPASPLDSTFYRSLLEDDMDGLDVA 1020
QY 1015 BEYLVPQGGFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEAAPSLAPSEG 1074
Db 1021 BEYLVPQGGFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEAAPSLAPSEG 1080
QY 1075 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPOPEYV 1134
Db 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPOPEYV 1140
QY 1135 NQPDVRQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVYKDVFAFGGAVENPEYLTPO 1194

Db	1141	NOQVDRPQPSREGPLPAAPAGATLPRKTLSPGKNGVWVDVFAFGAVENPEYLTPO	1200
Qy	1195	GGAAPOPHPPAFSPAFDNLXYWDQDPPRGAPPSTFKGTPTAENPEYLGLDVVP	1249
Db	1201	GGNAPOPHPPAFSPAFDNLXYWDQDPPRGAPPSTFKGTPTAENPEYLGLDVVP	1255
RESULT 12			
ID	AAE20479	AAE20479	
XX	AAE20479	standard; Protein; 1255 AA.	
AC	AAE20479;		
XX	01-JUL-2002	(first entry)	
DT	Human Her-2/neu protein.		
DE	Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;		
XX	human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.		
KW	Homo sapiens.		
XX	OS		
XX	Key	Location/Qualifiers	
FT	Region	1021..1030	
FT	/note=	"Naturally processed HLA-B44-restricted epitope"	
XX	WO200214503-A2.		
PN	21-FEB-2002.		
PD	14-AUG-2001; 2001WO-US41733.		
XX	14-AUG-2000; 2000US-225152P.		
XX	28-SEP-2000; 2000US-236428P.		
PR	21-FEB-2001; 2001US-270520P.		
XX	(CORI-) CORIXA CORP.		
PA	Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;		
XX	McNeill PD, Vedvick TS;		
PI	WPI; 2002-280758/32.		
XX	N-PSDB; AAD32743.		
DR	Novel isolated Her-2/Neu polypeptide composition useful for therapy,		
XX	prevention and diagnosis of cancer, preferably breast cancer		
PT	Disclosure; Page 114-117; 129pp; English.		
XX	The invention relates to an isolated Her-2/Neu polypeptide composition		
CC	effective for eliciting an immune response. The invention is useful for		
CC	eliciting an immune response in a patient, where the patient is human		
CC	leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.		
CC	The composition is useful for the therapy and diagnosis of cancer,		
CC	preferably breast cancer, in pharmaceutical compositions, e.g., vaccine		
CC	and other compositions for the diagnosis, prevention and treatment of		
CC	human malignancies, for stimulating and/or expanding T cells specific for		
CC	Her-2/Neu polypeptide and for inhibiting the development of cancer in a		
CC	patient. The invention is useful for stimulating a T cell response in a		
CC	human patient, as probe or primer for nucleic acid hybridisation, to		
CC	selectively form duplex molecules with complementary stretches of the		
CC	entire Her-2/Neu gene or gene fragments of interest, to isolate a full		
CC	length gene from a suitable library, and to direct expression of a		
CC	polypeptide in appropriate host cells. The composition is useful in		
CC	prophylactic or therapeutic applications and for the treatment of cancer,		
CC	preferably for the immunotherapy of breast cancer and other Her-2/Neu-		
CC	associated malignancies. The invention is useful in gene therapy. The		
CC	present sequence is human Her-2/neu protein.		
XX	Sequence	1255 AA;	
SQ	Query Match	98.5%; Score 6672; DB 23; Length 1255;	

Best Local Similarity		98.2%;	Pred. No. 0;		
Matches 1232; Conservative		5;	Mismatches	12;	Indels
		6;	Gaps		
Qy	1	MELAAALCWGGLLLALLPPGAAS	TOVCT-----OYIKANSKF	IGITTELLYQCQVQGNL	54
Db	1	MELAAALCWGGLLLALLPPGAAS	TOVCTGTGTDKMLRLPASPETH	LDMLRHLYQCQVQGNL	60
Qy	55	ELTYLPTNASLSFLQDIOEQVGY	VLIAHNRQVPLQRLRIVRG	TQLFEDNYALVDNG	114
Db	61	ELTYLPTNASLSFLQDIOEQVGY	VLIAHNRQVPLQRLRIVRG	TQLFEDNYALVDNG	120
Qy	115	DPLNNTTPTVGTASPGGLRELQ	LSLTELKGGVLIQRPOLCYQ	DTILWKDIFHKNNOLA	174
Db	121	DPLNNTTPTVGTASPGGLRELQ	LSLTELKGGVLIQRPOLCYQ	DTILWKDIFHKNNOLA	180
Qy	175	LTLIDTNRSRACHPCSPMKSGR	SGESSEDCQSLTRTVACG	GCARCKGPLTDCCHEOC	234
Db	181	LTLIDTNRSRACHPCSPMKSGR	SGESSEDCQSLTRTVACG	GCARCKGPLTDCCHEOC	240
Qy	235	AAGCTGPKHSDCLACLFHNSG	ICELHCPALVTYNTDTFES	MNPPEGRYTFGASCVTACP	294
Db	241	AAGCTGPKHSDCLACLFHNSG	ICELHCPALVTYNTDTFES	MNPPEGRYTFGASCVTACP	300
Qy	295	YNYLSTDVSGSCTLVCPHLNQ	EVTAEDGTQCEKSKPCAR	VCYGLGMEHLREVRVTSAN	354
Db	301	YNYLSTDVSGSCTLVCPHLNQ	EVTAEDGTQCEKSKPCAR	VCYGLGMEHLREVRVTSAN	360
Qy	355	IQEPAGCKKI7GSLAFLPES	FDGDPASNTAPLOPEOLQ	VFETLEITGYLYISAMPD	SLP 414
Db	361	IQEPAGCKKI7GSLAFLPES	FDGDPASNTAPLOPEOLQ	VFETLEITGYLYISAMPD	SLP 420
Qy	415	DLVSFQNLQVIRGRILHNGAY	SLTLQGLIGISWLGRLSR	RELGSGLALIHNTLHLCF	VHTV 474
Db	421	DLVSFQNLQVIRGRILHNGAY	SLTLQGLIGISWLGRLSR	RELGSGLALIHNTLHLCF	VHTV 480
Qy	475	PWDQLFRNPQALLHTANRPE	DECVGEGIACHQIARGHC	WGPGTQCVCNSQFIRG	QEC 534
Db	481	PWDQLFRNPQALLHTANRPE	DECVGEGIACHQIARGHC	WGPGTQCVCNSQFIRG	QEC 540
Qy	535	VEECRVLOGLPREYVNAHCL	PCHEPCOPNGSVTCFGE	ADQCACAHYKDPPECV	ARC 594
Db	541	VEECRVLOGLPREYVNAHCL	PCHEPCOPNGSVTCFGE	ADQCACAHYKDPPECV	ARC 600
Qy	595	PSGVKPDLSYMPKFPDEE	GACQPCINCTHSCVDL	DDKGPAPQORASPLTS	ISAVVG 654
Db	601	PSGVKPDLSYMPKFPDEE	GACQPCINCTHSCVDL	DDKGPAPQORASPLTS	ISAVVG 660
Qy	655	ILLVVVLGVVFGILIKRQO	KIRKYTMRRLLQETEL	VEPLTPSGAMPNOAQM	RILKETEL 714
Db	661	ILLVVVLGVVFGILIKRQO	KIRKYTMRRLLQETEL	VEPLTPSGAMPNOAQM	RILKETEL 720
Qy	715	RKVKVLGSGAGTGYKGIW	IPDGENVKIPVAIKVLR	ENTSPKANKEIIDEAY	VMAVGQSP 774
Db	721	RKVKVLGSGAGTGYKGIW	IPDGENVKIPVAIKVLR	ENTSPKANKEIIDEAY	VMAVGQSP 780
Qy	775	VYSRLLGICLTSTVOLVTQ	LMPIYGLLDHVRNRLGS	QDILLNMCQIAGKMS	YLEDVR 834
Db	781	VYSRLLGICLTSTVOLVTQ	LMPIYGLLDHVRNRLGS	QDILLNMCQIAGKMS	YLEDVR 840
Qy	835	LVHRDLAARNVLKSPNHV	KITDFGLARLLDIDETE	YHADGKVPKKNWAL	ESILRRRPT 894
Db	841	LVHRDLAARNVLKSPNHV	KITDFGLARLLDIDETE	YHADGKVPKKNWAL	ESILRRRPT 900
Qy	895	HQSDVMSGVTVWELMTG	AKPYDIPAREIPDLLEK	GERLPPOPICITD	IVYMINVKCM 954
Db	901	HQSDVMSGVTVWELMTG	AKPYDIPAREIPDLLEK	GERLPPOPICITD	IVYMINVKCM 960
Qy	955	IDSECRPRFRLVSEFS	RMARDPQRFVVIQNE	DLPASPLDSTFYR	SLEDDDDMDGLVDA 1014
Db	961	IDSECRPRFRLVSEFS	RMARDPQRFVVIQNE	DLPASPLDSTFYR	SLEDDDDMDGLVDA 1020
Qy	1015	EYLVPOQGFCCPDPA	GAGGMVHRRHSSSTR	SGGDLTLGLSEF	SEEPASPLAPSEG 1074

Db 1021 EEYLVPPQGFCDPAPAGAGMHHRRSSSTRSGGDLTLGLEPSEEBAPRSLAPSRG 1080
QY 1075 AGSDVFDGLMGAAKGLQSLTHDPSPLOQYSEDTPLPSPSTDCGYVAPLTCSPQPEYV 1134
Db 1081 AGSDVFDGLMGAAKGLQSLTHDPSPLOQYSEDTPLPSPSTDCGYVAPLTCSPQPEYV 1140
QY 1135 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPGKNGVVDVFAFGGAVENPEYLTTPQ 1194
Db 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPGKNGVVDVFAFGGAVENPEYLTTPQ 1200
QY 1195 GGAAPQHPHPPAFSPAFDNLVYWDQPPRGPAPPSFTFKGTPTAENPEYLGLDVVP 1249
Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPRGPAPPSFTFKGTPTAENPEYLGLDVVP 1255

RESULT 13

AAW51143
ID AAW51143 standard; Protein; 1255 AA.

XX AAW51143;

DT 17-JUN-2002 (first entry)

XX Human Her-2/neu oncogene-encoded p185 glycoprotein.

XX Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;

KW tyrosine kinase; receptor; c-erbB2; gene therapy.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Domain 1..653

FT /note= "extracellular domain"

FT Domain 676..1255

FT /note= "intracellular domain"

FT Domain 930..1255

FT /note= "phosphorylation domain"

XX WO200212341-A2.

XX 14-FEB-2002.

XX 03-AUG-2001; 2001WO-US24283.

XX 03-AUG-2000; 2000US-0632507.

XX (CORI-) CORIXA CORP.

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Cheever MA, Gheysen D;

XX WPI; 2002-241743/29.

DR N-PSDB; ABA92250.

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting

PT or enhancing an immune response to the protein, has Her-2/neu

PT extracellular domain fused to Her-2/neu intracellular or

PT phosphorylation domain -

XX Claim 68; Fig 7; 141pp; English.

XX The present sequence is that of human Her-2/neu (p185 glycoprotein
CC or c-erbB2), an oncogenic self-protein and target for anti-cancer
CC vaccines. The Her-2/neu gene is amplified and target for anti-cancer
CC in a variety of cancers, including breast, ovarian, colon, lung and
CC prostate cancer. Her-2/neu is a member of the tyrosine kinase
CC family of receptor-like glycoproteins. It comprises an extracellular
CC domain with homology to the epidermal growth factor receptor
CC (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
CC intracellular domain that also shows homology to EGFR. Its
CC overexpression correlates with a poor prognosis in breast and
CC ovarian cancers. The invention provides Her-2/neu fusion
CC proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In
CC preferred fusion proteins, the extracellular domain of a Her-2/neu
CC protein is fused to a Her-2/neu intracellular domain or
CC phosphorylation domain (or its DeltapD fragment). An immune
CC response to Her-2/neu protein is elicited or enhanced by
CC administering the fusion protein in the form of a vaccine, or by
CC transfecting cells of an animal ex vivo with a nucleic acid
CC encoding the fusion protein, and delivering the transfected cells
CC to the animal. The fusion proteins, nucleic acids, and isolated
CC specific T-cells are useful for inhibiting the development of a
CC cancer, especially breast, ovarian, colon, lung or prostate cancer
CC in a patient. T cells that specifically react with a Her-2/neu
CC fusion protein can be used to remove tumour cells from a sample in
CC order to inhibit the development of cancer in a patient.

XX Sequence 1255 AA;

Query Match 98.5%; Score 6672; DB 23; Length 1255;

Best Local Similarity 98.2%; Pred. No. 0;

Matches 1232; Conservative 5; Mismatches 12; Indels 6; Gaps 1;

QY 1 MELAALCRWGLLLALLPFGAASSTQVCT-----QYIKANSKFIGITELLYQCCQVQGNL 54

Db 1 MELAALCRWGLLLALLPFGAASSTQVCTGCTDMKRLPASPETHLDMLRHLQCCQVQGNL 60

QY 55 ELTYLPTNASLSFLQDIOEVQGYVLIAHNOVROVPLQRLIRVRGTQLFEDNYVALVDNG 114

Db 61 ELTYLPTNASLSFLQDIOEVQGYVLIAHNOVROVPLQRLIRVRGTQLFEDNYVALVDNG 120

QY 115 DPLNTTPTVGTASPGGLRELQRLSLEILKGVLIQPNPOLCYQDTILWKDIFHKNNQLA 174

Db 121 DPLNTTPTVGTASPGGLRELQRLSLEILKGVLIQPNPOLCYQDTILWKDIFHKNNQLA 180

QY 175 LTLIDTNRSRACHPCSPMKSGRSGESSEDCQSLTRTVCGAGCARCKGPLTDCHEQC 234

Db 181 LTLIDTNRSRACHPCSPMKSGRSGESSEDCQSLTRTVCGAGCARCKGPLTDCHEQC 240

QY 235 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGYTFGASCVTACP 294

Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGYTFGASCVTACP 300

QY 295 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVTSAN 354

Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVTSAN 360

QY 355 IOEFAGCKKIFGSLAFIPESFDGDPASNTAPLOPQLOVFETLEITGYLIYSAMPDSL 414

Db 361 IOEFAGCKKIFGSLAFIPESFDGDPASNTAPLOPQLOVFETLEITGYLIYSAMPDSL 420

QY 415 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGRLSRELGSGLALHNNTHLCFVHTV 474

Db 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGRLSRELGSGLALHNNTHLCFVHTV 480

QY 475 PWDQFRNPHQALLHTANRPEDECYGEGLACHQLCARGHCWGPPTQCVCNCSQFLRGQBC 534

Db 481 PWDQFRNPHQALLHTANRPEDECYGEGLACHQLCARGHCWGPPTQCVCNCSQFLRGQBC 540

QY 535 VEECRVLOGLPREYVYVNAHCLPCHPECPONGSVTCFGEADQCACAHYKDPFPCVARC 594

Db 541 VEECRVLOGLPREYVYVNAHCLPCHPECPONGSVTCFGEADQCACAHYKDPFPCVARC 600

QY 595 PSGVKPDI-SYMPIWKFPPDEGACQPCPINCSTHSCVDLDDKGPAPQASPLTSIVSAVVG 654

Db 601 PSGVKPDI-SYMPIWKFPPDEGACQPCPINCSTHSCVDLDDKGPAPQASPLTSIVSAVVG 660

QY 655 ILLVVVLGVFGILIKRQKQIRKYTMRLRLLQETELVEPLTPSGAMPNQAQMRILKETEL 714

Db 661 ILLVVVLGVFGILIKRQKQIRKYTMRLRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

QY 715 RKVKVLGSGAGTGYVKGWIIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVSP 774

Db 721 RKVKVLGSGAGTGYVKGWIIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVSP 780

QY 775 YVSRLLGICLTSTVQLVTLMPYGCILLDHVRENRLGSGDLLNWCQIAGKMSYLEDVR 834
 DB 781 YVSRLLGICLTSTVQLVTLMPYGCILLDHVRENRLGSGDLLNWCQIAGKMSYLEDVR 840
 QY 835 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNWALSILRRRT 894
 DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNWALSILRRRT 900
 QY 895 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYIMVKWM 954
 DB 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYIMVKWM 960
 QY 955 IDSECHPRRELVSERMRARPQRFVWQNEIDLGPASPLDSTFYRSILDEDDMGDLVDA 1014
 DB 961 IDSECHPRRELVSERMRARPQRFVWQNEIDLGPASPLDSTFYRSILDEDDMGDLVDA 1020
 QY 1015 EYLVLPQQGFPCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEERAPSPAPSEG 1074
 DB 1021 EYLVLPQQGFPCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEERAPSPAPSEG 1080
 QY 1075 AGSDVFDGLGMAAGKLSLPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1134
 DB 1081 AGSDVFDGLGMAAGKLSLPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 QY 1135 NQPDVRPQPPSPREGPLPAARAGATLERAKTILSPGKNGVWVDVFAFGAVENPEYLTQ 1194
 DB 1141 NQPDVRPQPPSPREGPLPAARAGATLERAKTILSPGKNGVWVDVFAFGAVENPEYLTQ 1200
 QY 1195 GGAAPOPHPPAFSPAFDNLVYWDQPPERGAAPPSTFKGTPTAENPEYLGLDVVPV 1249
 DB 1201 GGAAPOPHPPAFSPAFDNLVYWDQPPERGAAPPSTFKGTPTAENPEYLGLDVVPV 1255

RESULT 14

AAU77114

ID AAU77114 standard; Protein; 1255 AA.

XX AC AAU77114;

XX DT 05-JUN-2002 (first entry)

XX DE Human Her-2/neu polypeptide.

XX KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;

XX KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
 XX KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
 XX KW Hodgkin's lymphoma; T cell therapy.

XX OS Homo sapiens.

XX PN WO200213847-A2.

XX PD 21-FEB-2002.

XX PF 13-AUG-2001; 2001WO-US25408.

XX PR 14-AUG-2000; 2000US-0638280.

XX PR 28-SEP-2000; 2000US-0675904.

XX PA (CORI-) CORIXA CORP.

XX PI Gaiger A, Cheever MA, Hand-zimmermann S;

XX DR WPI; 2002-280741/32.

XX DR N-PSDB; ABK10730.

XX PT Inhibiting haematological malignancy development by administering
 PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
 PT encoding the polypeptide, or antigen presenting cells expressing the
 PT polypeptide -

XX PS Disclosure; Page 71-74; 74pp; English.

CC The invention relates to a method for inhibiting development of
 CC haematological malignancy in a patient by administering a polypeptide
 CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
 CC encoding the polypeptide. Antigen presenting cells that express the
 CC protein can also be administered. The sequences are used for inhibiting
 CC development of haematological malignancy such as acute myelogenous
 CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
 CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
 CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
 XX SQ Sequence 1255 AA;

Query Match 98.5%; Score 6672; DB 23; Length 1255;
 Best Local Similarity 98.2%; Pred. No. 0;
 Matches 1232; Conservative 5; Mismatches 12; Indels 6; Gaps 1;

QY 1 MELAALCRWGLLLALLPPGAASCTVCT-----QYIKANSKFIGITELLYQCCVQVQNL 54
 DB 1 MELAALCRWGLLLALLPPGAASCTVCTGDMKRLPASPETHLDMRLHUYQCCVQVQNL 60
 QY 55 ELTYLPTNASLSFLQDIOEVQGYVLIHNOVRQVPLQRLIRIVRGTLQDFEDNYALAVLNG 114
 DB 61 ELTYLPTNASLSFLQDIOEVQGYVLIHNOVRQVPLQRLIRIVRGTLQDFEDNYALAVLNG 120
 QY 115 DPLNNTTPTVGTASPGGLRELQRLSITELKGGVLIQRPOLCYQDTILWKDIFHKNNQLA 174
 DB 121 DPLNNTTPTVGTASPGGLRELQRLSITELKGGVLIQRPOLCYQDTILWKDIFHKNNQLA 180
 QY 175 LTLIDTNSRACHPCSPCKSGRSGESSEDCOSLTRTVCAGGCARCKGPTDCCHQEC 234
 DB 181 LTLIDTNSRACHPCSPCKSGRSGESSEDCOSLTRTVCAGGCARCKGPTDCCHQEC 240
 QY 235 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 294
 DB 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
 QY 295 YNYLSTDVSGCTLVCPLNQNEVTAEDGTQRCCKSKPCARVCYGLGMBHLREVRVTSAN 354
 DB 301 YNYLSTDVSGCTLVCPLNQNEVTAEDGTQRCCKSKPCARVCYGLGMBHLREVRVTSAN 360
 QY 355 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFEETLEETIGLYISAMPDSL 414
 DB 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFEETLEETIGLYISAMPDSL 420
 QY 415 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLRSRLGSLGALIHNNHLCFVHTV 474
 DB 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLRSRLGSLGALIHNNHLCFVHTV 480
 QY 475 PWDQLFRNPHQALLHTANRPEDECYGEGLACHQLCARGHCWGPPTQCVCNCSQFLRGQEC 534
 DB 481 PWDQLFRNPHQALLHTANRPEDECYGEGLACHQLCARGHCWGPPTQCVCNCSQFLRGQEC 540
 QY 535 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFCVARC 594
 DB 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFCVARC 600
 QY 595 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCSTHSCVDLDDKGPAPQORASPLTSIVSAVVG 654
 DB 601 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCSTHSCVDLDDKGPAPQORASPLTSIVSAVVG 660
 QY 655 ILLVVVLGVW7GILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 714
 DB 661 ILLVVVLGVW7GILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
 QY 715 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKIILDEAYVMAGVSP 774
 DB 721 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKIILDEAYVMAGVSP 780
 QY 775 YVSRLLGICLTSTVQLVTLMPYGCILLDHVRENRLGSGDLLNWCQIAGKMSYLEDVR 834
 DB 781 YVSRLLGICLTSTVQLVTLMPYGCILLDHVRENRLGSGDLLNWCQIAGKMSYLEDVR 840
 QY 835 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNWALSILRRRT 894

Db 841 LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALESIILRRFT 900
 QY 895 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYMWKWM 954
 Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYMWKWM 960
 QY 955 IDSECRPRFRELVSFESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDMDGLVDA 1014
 Db 961 IDSECRPRFRELVSFESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDMDGLVDA 1020
 QY 1015 EYLVPOQGFCCPDPAAGAGMWHHRSSSTRSGGDLTLGLEPSEEBAPRPLAPSEG 1074
 Db 1021 EYLVPOQGFCCPDPAAGAGMWHHRSSSTRSGGDLTLGLEPSEEBAPRPLAPSEG 1080
 QY 1075 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1134
 Db 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 QY 1135 NOPDVRPQPPSPREGPLPAARPAATLERAKTILSPGKNGVVDVFAFGAVENPEYLTPO 1194
 Db 1141 NOPDVRPQPPSPREGPLPAARPAATLERAKTILSPGKNGVVDVFAFGAVENPEYLTPO 1200
 QY 1195 GGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGDLVPV 1249
 Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGDLVPV 1255

RESULT 15

AAR39568

ID AAR39568 standard; Protein; 1433 AA.

XX AAR39568;

XX

DT 07-FEB-1994 (first entry)

DE Sequence of c-erbB-2 tumour antigen.

KW Tumour antigen; c-erbB-2; glycoprotein.

XX Homo sapiens.

XX

XX

XX

XX

XX

PF 05-FEB-1993; 93WO-US01055.

PR 06-FEB-1992; 92US-0831967.

PA (CETU) CETUS ONCOLOGY CORP.

PA (CREA-) CREATIVE BIOMOLECULES INC.

XX

PI Houston LL, Huston JS, Oppermann H, Ring DB;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Matches 1225; Conservative 7; Mismatches 17; Indels 6; Gaps 1;
 QY 1 MELAALCWGILLALLPPCASTOYCT-----QYIKANSKEFIGITTELYLQCOVQGNL 54
 Db 1 MELAALCWGILLALLPPGAASCTOYCTGDMKRLRUPASPTHLDMLRLHYQCOVQGNL 60
 QY 55 ELTYLPTNASISFLQDIOEVQYVLIHNRQVPLQRLRIVRGTLQFEDNYALAVLNG 114
 Db 61 ELTYLPTNASISFLQDIOEVQYVLIHNRQVPLQRLRIVRGTLQFEDNYALAVLNG 120
 QY 115 DPLNNTTPTVGTASPGGLRELQRLSLEILKGGVLIQORNQOLCYQDTILWKDIFHKNNQLA 174
 Db 121 DPLNNTTPTVGTASPGGLRELQRLSLEILKGGVLIQORNQOLCYQDTILWKDIFHKNNQLA 180
 QY 175 LTLIDTNSRACHPCSPMKSGRCWGESSEDCOSLTRITVCAGGCARCKGLPTDCHEOC 234
 Db 181 LTLIDTNSRACHPCSPMKSGRCWGESSEDCOSLTRITVCAGGCARCKGLPTDCHEOC 240
 QY 235 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGYTFCASCVTACP 294
 Db 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGYTFCASCVTACP 300
 QY 295 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 354
 Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 360
 QY 355 IOEPAGCKKIFGSLAFLPESPDGPASNTAPLOPELOVFFETLEETITGLYIISANPDSL 414
 Db 361 IOEPAGCKKIFGSLAFLPESPDGPASNTAPLOPELOVFFETLEETITGLYIISANPDSL 420
 QY 415 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLRSRELGSGLALHNNHLCFVHTV 474
 Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLRSRELGSGLALHNNHLCFVHTV 480
 QY 475 PWDQLFRNPHQALLHTANRPEDECYVGEGLACHQLCARGHCWGPGTQCVCNSQFIRGQEC 534
 Db 481 PWDQLFRNPHQALLHTANRPEDECYVGEGLACHQLCARGHCWGPGTQCVCNSQFIRGQEC 540
 QY 535 VEECHVLOGLPREYNASHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFCVARC 594
 Db 541 VEECHVLOGLPREYNASHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFCVARC 600
 QY 595 PSGVKPDLSPYMPKPFDEBACQPCPINCTSHSCVDLDDKGPAPQASPLTISVAVVG 654
 Db 601 PSGVKPDLSPYMPKPFDEBACQPCPINCTSHSCVDLDDKGPAPQASPLTISVAVVG 660
 QY 655 ILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNOAQRILKETEL 714
 Db 661 ILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNOAQRILKETEL 720
 QY 715 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAYVMAGVSP 774
 Db 721 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAYVMAGVSP 780
 QY 775 YVSRLLGICLTSTVOLVTLQMPYGLLDHVRNRLGSDLLNWCMIKAGMSYLEDVR 834
 Db 781 YVSRLLGICLTSTVOLVTLQMPYGLLDHVRNRLGSDLLNWCMIKAGMSYLEDVR 840
 QY 835 LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALESIILRRFT 894
 Db 841 LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALESIILRRFT 900
 QY 895 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYMWKWM 954
 Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYMWKWM 960
 QY 955 IDSECRPRFRELVSFESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDMDGLVDA 1014
 Db 961 IDSECRPRFRELVSFESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDMDGLVDA 1020
 QY 1015 EYLVPOQGFCCPDPAAGAGMWHHRSSSTRSGGDLTLGLEPSEEBAPRPLAPSEG 1074
 Db 1021 EYLVPOQGFCCPDPAAGAGMWHHRSSSTRSGGDLTLGLEPSEEBAPRPLAPSEG 1080

New single chain Fv polypeptide binding to C-erbB-2 tumour antigen - for imaging or treating breast or ovarian cancer etc.
 Disclosure; pages 48-54; 87pp; English.

c-erbB-2 refers to a protein antigen expressed on the surface of tumour cells. such as breast and ovarian tumour cells, which is an approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39568 represents the location of a stop codon in AAQ46083.

Query Match 97.9%; Score 6629; DB 14; Length 1433;
 Best Local Similarity 97.6%; Pred. No. 0;

QY	1075	AGSDVFDGLGMAAKGLQSLPTHDPSPLOQYSEDTVPLPSETDGYVAPLTCSPQPEYV	1134
Db	1081	AGSDVFDGLGMAAKGLQSLPTHDPSPLOQYSEDTVPLPSETDGYVAPLTCSPQPEYV	1140
QY	1135	NOPDVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKDVFAFGGAVENPEYLTTPQ	1194
Db	1141	NOPDVRPQPPSPREGPLPAARPAGATLERPKTILSPGKNGVVKDVFAFGGAVENPEYLTTPQ	1200
QY	1195	GGAAPQPHPPPAFSPAFDNLYYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVPV	1249
Db	1201	GGAAPQPHPPPAFSPAFDNLYYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVPV	1255

Search completed: July 22, 2003, 08:40:06
Job time : 42.7958 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 22.9062 Seconds
(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-59-73-12
Perfect score: 6809
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGDPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6729	98.8	1255	1 A24571	protein-tyrosine k
2	5919	86.9	1260	1 TVRTNU	protein-tyrosine k
3	5918.5	86.9	1254	2 I48161	p-185 precursor -
4	3117	45.8	1210	1 GQHUE	epidermal growth f
5	3093	45.4	1210	2 A53183	epidermal growth f
6	3085.5	45.3	1223	1 TVCHLV	epidermal growth f
7	2950.5	43.3	1308	2 A47253	epidermal growth f
8	2653	39.0	1166	1 S06142	protein-tyrosine k
9	2389.5	35.1	1342	2 A36223	kinase-related tra
10	2316.5	34.0	1339	2 JC4387	epidermal growth f
11	1786.5	25.9	698	1 TVFVLV	protein-tyrosine k
12	1703	25.0	604	1 TVYUHV	protein-tyrosine k
13	1654.5	24.3	1330	1 GQFFE	epidermal growth f
14	1647	24.2	544	2 S35745	protein-tyrosine k
15	1640	24.1	545	2 S00727	kinase-related tra
16	1623	23.8	540	2 B44776	protein-tyrosine k
17	1621	23.8	540	1 TVFVBB	protein-tyrosine k
18	1485	21.8	644	2 A36325	epidermal growth f
19	1275	18.7	1323	2 E88257	protein let-23 (im
20	1275	18.7	1374	2 S70712	protein-tyrosine k
21	1187	17.4	1369	2 S70713	protein-tyrosine k
22	1146	16.8	1717	1 A45558	epidermal growth f
23	1117	16.4	527	2 A42032	epidermal growth f
24	955.5	14.0	843	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	protein-tyrosine k
26	754.5	11.1	311	2 S13808	protein-tyrosine k
27	742	10.9	1363	2 T43220	insulin-like growt
28	717	10.5	1382	1 INHUR	insulin receptor p
29	710	10.4	1383	2 A36080	insulin receptor p

30	709.5	10.4	1372	2 A34157	insulin receptor p
31	700	10.3	1607	2 T43212	insulin-like growt
32	688.5	10.1	1300	2 A36502	insulin receptor-r
33	676	9.9	1268	2 B36502	insulin receptor-r
34	672	9.9	1477	2 T18534	protein-tyrosine k
35	653	9.6	1367	1 IGHUR1	insulin-like growt
36	642	9.4	1371	2 A33837	insulin-like growt
37	635	9.3	1390	2 T30346	insulin receptor -
38	618.5	9.1	2148	1 A56081	insulin receptor (
39	613	9.0	2101	2 S57245	insulin receptor (
40	610	9.0	987	2 A54092	protein-tyrosine k
41	591.5	8.7	952	2 I50612	protein-tyrosine k
42	589.5	8.7	977	2 S49004	tyrosine kinase Mp
43	588	8.6	1114	1 S05582	protein-tyrosine k
44	586	8.6	987	2 I48652	mouse developmenta
45	586	8.6	1091	2 S33596	protein-tyrosine k

ALIGNMENTS

RESULT 1

A24571
N;Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein erbB;
C;Species: Homo sapiens (man)
C;Date: 25-Oct-1987 #sequence.revision 06-Dec-1996 #text change 11-Jun-1999
C;Accession: A24571; A25491; A44188; B44188; I59509; I57622
R;Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;
Nature 319, 230-234, 1986
A;Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth f.
A;Reference number: A24571; MUID:86118663; PMID:3003577
A;Accession: A24571
A;Molecule type: mRNA
A;Residues: 1-1255 <YAM>
R;Cross-references: GB:X03363; NID:G31197; PIDN:CAA27060.1; PID:G31198
R;Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A;Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epiderma
A;Reference number: A25491; MUID:86016729; PMID:2995967
A;Accession: A25491
A;Molecule type: DNA
A;Residues: 737-1031 <SEM>
A;Cross-references: GB:M11767; NID:G182163; PIDN:AAA35808.1; PID:G553282
R;Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg, P
Science 230, 1132-1139, 1985
A;Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromos
A;Reference number: A44188; MUID:86070181; PMID:2999974
A;Accession: A44188
A;Molecule type: DNA
A;Residues: 740-910 <COU1>
A;Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989
A;Accession: B44188
A;Molecule type: mRNA
A;Residues: 1-517, 'FALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>
A;Cross-references: GB:M11730; NID:G183986
R;King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 229, 974-976, 1985
A;Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
A;Reference number: I59509; MUID:85272597; PMID:2992089
A;Accession: I59509
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 832-909 <REX>
A;Cross-references: GB:L29395; NID:G459807; PIDN:AAA35809.1; PID:G459808
R;Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987
A;Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptional
A;Reference number: I57622; MUID:87286898; PMID:3039351
A;Accession: I57622
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-191 <TAL>

A;Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C;Genetics: GDB:ERBB2; NGL; NEU; HER-2
A;Gene: GDB:ERBB2; NGL; NEU; HER-2
A;Cross-references: GDB:120613; OMIM:164870
A;Map position: 17q21.1-17q21.1
A;Introns: 25/1; 75/3; 147/1; 883/3
A;Note: the list of introns is incomplete
C;Function:
A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
inase
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F:722-653/Domain: extracellular #status predicted <EXT>
F:770-304/Domain: EGF receptor extracellular domain repeat <EE1>
F:395-605/Domain: EGF receptor extracellular domain repeat <EE2>
F:654-675/Domain: transmembrane #status predicted <TM>
F:676-1255/Domain: intracellular #status predicted <INT>
F:718-983/Domain: protein kinase homology <KIN>
F:728-734/Region: protein kinase ATP-binding motif
F:168,124,187,259,530,571,629/Binding site: carboxydrate (Asn) (covalent) #status predicted
F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:753/Active site: Lys #status predicted
F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 98.8%; Score 6729; DB 1; Length 1255;
Best Local Similarity 98.8%; Pred. NO. 1.6e-268;
Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLPPGAASCTCTDMKRLPASPETHDMLRLHYOGCQVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASCTCTDMKRLPASPETHDMLRLHYOGCQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIEVQOYIKANSKFIGITELQRLRVGRGTQLPEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIEVQOYIAHNOVQVPLQRLRVGRGTQLPEDNYALAVLDNG 120
Qy 121 DPLNTTPTVGTASPGGLRELQRLSUTLILKGVLIQRPOLCYQDTILWKDIFHKNNOLA 180
Db 121 DPLNTTPTVGTASPGGLRELQRLSUTLILKGVLIQRPOLCYQDTILWKDIFHKNNOLA 180
Qy 181 LTLIDTNRSRACHPCSPKCSRGWSESDQSLTRVTCAGGACRCKGLPTDCCHQC 240
Db 181 LTLIDTNRSRACHPCSPKCSRGWSESDQSLTRVTCAGGACRCKGLPTDCCHQC 240
Qy 241 AAGCTGPKRHSCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300
Db 241 AAGCTGPKRHSCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300
Qy 301 YNYLSTVGSCTLVCLNHNQVETADGTQRCCKSPCARVCYGLGMHRLREVRAVTSAN 360
Db 301 YNYLSTVGSCTLVCLNHNQVETADGTQRCCKSPCARVCYGLGMHRLREVRAVTSAN 360
Qy 361 IOEFAGCKKIFGSLAFLPESDGPDPASNTAPLOEQLOVFELEITCYLYISAWPDSLP 420
Db 361 IOEFAGCKKIFGSLAFLPESDGPDPASNTAPLOEQLOVFELEITCYLYISAWPDSLP 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLSLTQGLIGISWGLRSRLRELGSGLALHNNHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLSLTQGLIGISWGLRSRLRELGSGLALHNNHLCFVHTV 480
Qy 481 PWDQFLRNPQALLHTANRPDECVGEGGLACHOLCARGHCWGPGTQCVCNCSQFLRGQC 540
Db 481 PWDQFLRNPQALLHTANRPDECVGEGGLACHOLCARGHCWGPGTQCVCNCSQFLRGQC 540
Qy 541 VEECKVLQGLPREYVNAHRLCPCHPECOPOGNSVTCFGEADQCVACAHYKDPFCVARC 600
Db 541 VEECKVLQGLPREYVNAHRLCPCHPECOPOGNSVTCFGEADQCVACAHYKDPFCVARC 600
Qy 601 PSGVKPDLSPYMPIKWPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTISAVVG 660
Db 601 PSGVKPDLSPYMPIKWPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTISAVVG 660

Db 601 PSGVKPDLSPYMPIKWPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTISAVVG 660
Qy 661 ILLVVLGVVFGILIKRROOKIRKYTMRRLQETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVVLGVVFGILIKRROOKIRKYTMRRLQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILD EAYVWAGVSP 780
Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILD EAYVWAGVSP 780
Qy 781 YVSRLLIGICLTSTVQLVTLMPYGCCLLDHVRNRLGRLGSDLLNWCNQIAGKMSYLEDVR 840
Db 781 YVSRLLIGICLTSTVQLVTLMPYGCCLLDHVRNRLGRLGSDLLNWCNQIAGKMSYLEDVR 840
Qy 841 LVHRDLAARNVLYKSPNHVKITDFGLARLDDIDETEHADGGKVPICKMALESILRRRET 900
Db 841 LVHRDLAARNVLYKSPNHVKITDFGLARLDDIDETEHADGGKVPICKMALESILRRRET 900
Qy 901 HQSDVMSYGYTVWELMTFGAKPYDGPAREIPDLEKGERLPPOPICTIDVYIMVAKCM 960
Db 901 HQSDVMSYGYTVWELMTFGAKPYDGPAREIPDLEKGERLPPOPICTIDVYIMVAKCM 960
Qy 961 IDSECRPRFRELVSFSEFARMARDPQRFVVIQNEDLGPASPLDSTFYRSLLED DDDMDGLVDA 1020
Db 961 IDSECRPRFRELVSFSEFARMARDPQRFVVIQNEDLGPASPLDSTFYRSLLED DDDMDGLVDA 1020
Qy 1021 EYVLVPOOGFCDDPAPGAGMVHHRSSSTRSGGDLTLGLEPSEEEAPRSPPLASEG 1080
Db 1021 EYVLVPOOGFCDDPAPGAGMVHHRSSSTRSGGDLTLGLEPSEEEAPRSPPLASEG 1080
Qy 1081 AGSDVDFGDLGMAAKQLSLPTHDPSPLOQYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVDFGDLGMAAKQLSLPTHDPSPLOQYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Qy 1141 NQPDVRPQPPSPREGPLPAAPAGATLERAKTILSPGKGVVQKVFAGFGAVENPEYLTPO 1200
Db 1141 NQPDVRPQPPSPREGPLPAAPAGATLERAKTILSPGKGVVQKVFAGFGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYGLDVPV 1255
Db 1201 GGAAPQHPHPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYGLDVPV 1255

RESULT 2
TVRTNU
Protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
C;Accession: A24562; A61204
R;Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A;Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.
A;Reference number: A24562; MUID:86118662; PMID:3945311
A;Accession: A24562
A;Molecule type: mRNA
A;Residues: 1-1260 <BAR>
A;Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746
R;Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen,
Carcinogenesis 12, 1975-1978, 1991
A;Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no m
2-thiazolyl]formamide or N-methyl-N-nitrosourea.
A;Reference number: A61204; MUID:92035293; PMID:1682063
A;Accession: A61204
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 637-663, 'V', 665-702 <MAS>
A;Note: authors translated the codon GCA for residue 25 as Val
C;Genetics:
A;Gene: neu
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

F;658-680/Domain: transmembrane #status predicted <TM>
 F;723-988/Domain: protein kinase homology <KIN>
 F;731-739/Region: protein kinase ATP-binding motif
 F;71,191,263,535,576,634/Binding site: carboxylate (Asn) (covalent) #status predicted
 F;691/Binding site: phosphate (Thr) (covalent) #status predicted
 F;758/Active site: Lys #status predicted
 F;882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 86.9%; Score 5919; DB 1; Length 1260;
 Best Local Similarity 87.0%; Pred. No. 2.4e-235;
 Matches 1093; Conservative 49; Mismatches 113; Indels 2; Gaps 2;

Qy 1 MELAALCRWGLLLALLPPGAASSTQVCTDKMLRLPASPTHLDMLRHLHYGCGVQVGNL 60
 Db 4 MELAALCRWGLLLALLPPGIAAGTQVCTDKMLRLPASPTHLDMLRHLHYGCGVQVGNL 63
 Qy 61 ELTYLPTNASLSFLQDIQEVQYIKANSKFTIGTELQRLRVRGTQTFEDNYALAVLDNG 120
 Db 64 ELTYVPANASLSFLQDIQEVQYMLIAHNVKRVPLQRLRVRGTQTFEDNYALAVLDNR 123
 Qy 121 DPLNNTPTVT-GASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILKWDIFHKNNQL 179
 Db 124 DPQDNVAASLTPGRTPEGLRELQRLSLTEILKGGVLIQRNPOLCYQDMVLMKDFVRKNNQL 183
 Qy 180 ALTLIDNRSRACHPCSPMKGSCWGESSEDDCSLRTVTCAGGCARCKGRLPTDCCHEQ 239
 Db 184 APVDIDNRSRACHPCSPMKGSCWGESSEDDCSLRTVTCAGGCARCKGRLPTDCCHEQ 243
 Qy 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTAC 299
 Db 244 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTTC 303
 Qy 300 PYNLSTDVSGCTLVCLPHNOEVTAEADGTQRCCKSPCARVCYGLGMEHLRVRVTS 359
 Db 304 PYNLSTEVSGCTLVCLPHNOEVTAEADGTQRCCKSPCARVCYGLGMEHLRVRVTS 363
 Qy 360 NIQFACGKIFGLAFPLPSFDGDPASNTAPLOEQLOVPEETBEITYIYISAWPDSL 419
 Db 364 NVQFEDGCKIFGLAFPLPSFDGDPASNTAPLOEQLOVPEETBEITYIYISAWPDSL 423
 Qy 420 PDLVFQNLQVIRGRIILHNGAYSITLQGLGTSWGLRSLRELGSGLALIHNNHLCVFHT 479
 Db 424 RDLVFQNLQVIRGRIILHNGAYSITLQGLGHSGLRSLRELGSGLALIHNNHLCVFHT 483
 Qy 480 VPWDQFRNPQALLTANPEDE-CVGEGLACHQLCARGHCWGPPTQCVCNCSQFLRGQ 538
 Db 484 VPWDQFRNPQALLHSGNRFEEDLCVSSGLVCNLSAHCWGPPTQCVCNCSHFLRGQ 543
 Qy 539 ECVEECRVQLGPREYVNAHCLCPHECQPDQSGVTCFGEADQCACAHYKDPDFCVA 598
 Db 544 ECVEECRVKGLPREYVSDKRLCPHECQPDQSGVTCFGEADQCACAHYKDSSCVA 603
 Qy 599 RCPGSGVDPDLSYPIWPKFPDEEGACQPCINCTHSCVDLDDKGPASQASPLTSIVSAV 658
 Db 604 RCPGSGVDPDLSYPIWPKFPDEEGACQPCINCTHSCVDLDDKGPASQASPLTSIVATV 663
 Qy 659 VGILLVVVGVVGFILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAQMRILKET 718
 Db 664 EGVLLELILVVVGVVGFILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAQMRILKET 723
 Qy 719 ELRKVVLGSGAFQTVYKGIWPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGV 778
 Db 724 ELRKVVLGSGAFQTVYKGIWPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGV 783
 Qy 779 SPYVRLGLCLTSTVOLVTLMPYGCGLDHHVNRGRGLSGODLLNKCQIAKMSYLED 838
 Db 784 SPYVRLGLCLTSTVOLVTLMPYGCGLDHHVNRGRGLSGODLLNKCQIAKMSYLED 843
 Qy 839 VRLVHRLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWALESLR 898
 Db 844 VRLVHRLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWALESLR 903
 Qy 899 FTHQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICTIDVYIMVKC 958

Db 904 FTHQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICTIDVYIMVKC 963
 Qy 959 WMTDSECRPHRELVSSEFSRMDPQRFVVIQNEIDLPASPLDSTFVRSLEDDMDGLV 1018
 Db 964 WMTDSECRPHRELVSSEFSRMDPQRFVVIQNEIDLPASPLDSTFVRSLEDDMDGLV 1023
 Qy 1019 DAEYLVPOQGFPPDPAPGAGMVRHRRSSSTRSGGDLTLGLEPSESEAPRSLAPS 1078
 Db 1024 DAEYLVPOQGFPPDPAPGAGMVRHRRSSSTRSGGDLTLGLEPSESEAPRSLAPS 1083
 Qy 1079 EGAGSDVFDGDLGWAAGKLOSLPHTDPSLQRYSEDPVPLPSETDGYVAPLTCSPQE 1138
 Db 1084 EGAGSDVFDGDLGWAAGKLOSLPHTDPSLQRYSEDPVPLPSETDGYVAPLTCSPQE 1143
 Qy 1139 YVQNPQVRPQPPPREGLPAAPAGATLERAKTLPCKNGVVKDVFAGGAVENPEVLT 1198
 Db 1144 YVQNPQVRPQPPPREGLPAAPAGATLERAKTLPCKNGVVKDVFAGGAVENPEVLT 1203
 Qy 1199 PQGGAAPQPPPPAFSPAFDNLVYWDQPPPERGAPSTFKGTPTAENPEYLGDDVPV 1255
 Db 1204 PREGTASPPHPSAFSPAFDNLVYWDQNSSEQGPSPPSNFEPTTAENPEYLGDDVPV 1260

RESULT 3
 I48161
 P-185 precursor - golden hamster
 C;Species: Mesocricetus auratus (golden hamster)
 C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
 C;Accession: I48161
 R;Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa, T.
 A;Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
 A;Reference number: I48151; MUID:94193007; PMID:7908275
 A;Accession: I48161
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-1254 <RES>
 A;Cross-references: GB:D16295; NID:G493236; PIDN:BAA03801.1; PID:G747595
 C;Genetics:
 A;Gene: neu
 C;Superfamily: epidermal growth factor receptor; protein kinase homology
 C;Keywords: ATP
 F;718-983/Domain: protein kinase homology <KIN>
 F;726-734/Region: protein kinase ATP-binding motif

Query Match 86.9%; Score 5918.5; DB 2; Length 1254;
 Best Local Similarity 86.8%; Pred. No. 2.5e-235;
 Matches 1089; Conservative 58; Mismatches 107; Indels 1; Gaps 1;

Qy 1 MELAALCRWGLLLALLPPGAASSTQVCTDKMLRLPASPTHLDMLRHLHYGCGVQVGNL 60
 Db 1 MELAALCRWGLLLALLPPGASSTQVCTDKMLRLPASPTHLDMLRHLHYGCGVQVGNL 60
 Qy 61 ELTYLPTNASLSFLQDIQEVQYIKANSKFTIGTELQRLRVRGTQTFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLSFLQDIQEVQYIKANSKFTIGTELQRLRVRGTQTFEDNYALAVLDNR 120
 Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILKWDIFHKNNQL 180
 Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILKWDIFHKNNQL 180
 Qy 181 LTLDNRSRACHPCSPMKGSCWGESSEDDCSLRTVTCAGGCARCKGRLPTDCCHEQC 240
 Db 181 PVDIDNRSRACHPCSPMKGSCWGESSEDDCSLRTVTCAGGCARCKGRLPTDCCHEQC 240
 Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTAC 300
 Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTTC 300
 Qy 301 PYNLSTDVSGCTLVCLPHNOEVTAEADGTQRCCKSPCARVCYGLGMEHLRVRVTSAN 360
 Db 301 PYNLSTEVSGCTLVCLPHNOEVTAEADGTQRCCKSPCARVCYGLGMEHLRVRVTSAN 360

Qy 361 IOEFAGCKIFGSLAFIPESFDGDPASNTAPLOEQLOVFTLEBEITCYLYISAWPDSLP 420
Db 361 IOEFAGCKIFGSLAFIPESFDGDPSSGIAPLTPEQLQVFTLEBEITCYLYISAWPDSLH 420
Qy 421 DLSVFQNLQVTRGRLHNGAYSLTLQIGISWGLRLSRLRELGSLALHHTHNLCHFVHTV 480
Db 421 DLSVFQNLQVTRGRLHNGAYSLTLQIGISWGLRLSRLRELGSLALHHTHNLCHFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQICARGHCWGPQPTQVCNCSQFLRGQEC 540
Db 481 PWDQLFRNPHQALLHSGNPSSEECGLKDFACYPICAHGHCWGPQPTQVCNCSHFLRGQEC 540
Qy 541 VEECRVLOGLPREYNABHCLPCHPECOPOGNSVTCFPEADQCVACAHYKDPFCVARC 600
Db 541 VKECRVWGLPREYNGKHCLPCHPECOPONSTETCTGSEADQCTACPHYKDSFFCVARC 600
Qy 601 PSGVKPDLSPYMPIWKFPEDEGACOPCINCTHSCVDLDKGCFAEQRASPLTSTVSAVVG 660
Db 601 PSGVKPDLSPYMPIWKFPEDEGACOPCINCTHSCVDLDKGCFAEQRASPLTSTVSATVVG 660
Qy 661 ILLVVVLGVGILLIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLFLVIGVWVGIILIKRRQKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
Qy 721 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVWAGVSP 780
Db 721 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVWAGLSP 780
Qy 781 YVSRLLGCLTSTVOLVTQMLPYGCLLDHVRNRLGRLSQDLLNWCMIAGKMSYLEDVR 840
Db 781 YVSRLLGCLTSTVOLVTQMLPYGCLLDHVRNRLGRLSQDLLNWCMIAGKMSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPKIKMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPKIKMALESILRRFT 900
Qy 901 HOSDWSVGVTVWELMTGAKPYGIPAREIPDLEKGERLPPOPICTIDVYMWVKCWM 960
Db 901 HOSDWSVGVTVWELMTGAKPYGIPAREIPDLEKGERLPPOPICTIDVYMWVKCWM 960
Qy 961 IDSECRPFRELVEFSSMARDPORFVVIQNEIDLGPSPLDSTFVRSILLEDDMGDLVDA 1020
Db 961 IDSECRPFRELVEFSSMARDPORFVVIQNEIDLGPSPLDSTFVRSILLEDDMGDLVDA 1020
Qy 1021 BEYLVPOQGFPCPDAPAGAGWVHRHRSSTRSGGDLTLGLPSEEEAPRSPAPSEG 1080
Db 1021 BEYLVPOQGFPCPDAPAGAGTAHRRHRSSTRSGGDLTLGLMPSGEPSPRSPAPSEG 1080
Qy 1081 AGSDVFDGLGKAGKGLQSLPHTDPSPLQYSEDPTVPLSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFEGLGKAGKGLQSPDLSPQYSEDPTVPLSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPOPSPREGPLPAAPAGATLRAKTLSPKNGVVKDVFAGGAVENPEYLTTPQ 1200
Db 1141 NOPEVRPOPPLTPEGPLPVPAGATLRLPKTLSPKNGVVKDVFAGGAVENPEYLVPR 1200
Qy 1201 GGAPAPHPHAFSPAFNLVYWDODPPERGAPSTFTKGTPTAENPEYGLDVPV 1255
Db 1201 GGSASQPH-PPALCPAFNLVYWDODPSERGSPPNFTGPTAENPEYGLDVPV 1254

RESULT 4
GQHUE
epidermal growth factor receptor precursor - human
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C:Species: Homo sapiens (man)
C:Date: 15-Nov-1984 #sequence revision 27-Nov-1985 #text change 11-Jun-1999
R:Ullrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.; Y
rg, P.H.
A:Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression of
Nature 309, 418-425, 1984

A:Reference number: A00641; MUID:84219729; PMID:6328312
A:Accession: A00641
A:Molecule type: mRNA
A:Residues: 1-1210 <ULL>
A:Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA25240.1; PID:g575924
A:Note: the authors translated the codon AAG for residue 540 as Asn
R:Ishii, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
A:Title: Characterization and sequence of the promoter region of the human epidermal growth factor receptor gene
A:Reference number: A25772; MUID:85270438; PMID:2991899
A:Accession: A25772
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-29 <ISH>
A:Cross-references: GB:M11234; NID:g181981; PIDN:AAA52370.1; PID:g553272
R:Haley, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ullrich, A.; Waterfield, M.
Oncogene Res. 1, 375-396, 1987
A:Title: The human EGF receptor gene: structure of the 110 kb locus and identification of the promoter
A:Reference number: S30024; MUID:88217333; PMID:3329716
A:Accession: S30024
A:Molecule type: DNA
A:Residues: 1-29 <HA2>
A:Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
R:Haley, J.D.; Waterfield, M.D.
J. Biol. Chem. 266, 1746-1753, 1991
A:Title: Contributory effects of de Novo transcription and premature transcript termination on the expression of the human epidermal growth factor receptor cDNA
A:Reference number: A38672; MUID:91107677; PMID:1988448
A:Accession: A38672
A:Molecule type: DNA
A:Residues: 1-29 <HAL>
A:Cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271
A:Experimental source: carcinoma cell line A431-7
R:Xu, Y.; Ishii, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.; Merlino, G.T.; Pastan, I.
Nature 309, 806-810, 1984
A:Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RNAs
A:Reference number: A00642; MUID:84245835; PMID:6330563
A:Accession: A00642
A:Molecule type: mRNA
A:Residues: 1-1087, 'KSVIQAV', 195, 'M', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-321
'798-799, 'TD', 802-811, 'R', 813-942 <XOY>
A:Experimental source: A431 human carcinoma cells, which have large numbers of EGF receptor
R:Lin, C.R.; Chen, W.S.; Krueger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verma, I.M.
Science 224, 843-848, 1984
A:Title: Expression cloning of human EGF receptor complementary DNA: gene amplification
A:Reference number: A43615; MUID:84196372; PMID:6326261
A:Accession: A43615
A:Molecule type: mRNA
A:Residues: 713-964 <LIN>
A:Experimental source: epidermoid carcinoma cell line A431
R:Simmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.
Biochem. Biophys. Res. Commun. 124, 125-132, 1984
A:Reference number: A23062; MUID:85046483; PMID:6093780
A:Accession: A23062
A:Molecule type: mRNA
A:Residues: 1028-1210 <SIM>
R:Weber, W.; Gill, G.N.; Speiss, J.
Science 224, 294-297, 1984
A:Reference number: A05281; MUID:84172183; PMID:6324343
A:Accession: A05281
A:Molecule type: protein
A:Residues: 25-30, 'S', 32-51, 454-467 <WEB>
R:Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.
J. Biol. Chem. 260, 5205-5208, 1985
A:Title: Identification of residues in the nucleotide binding site of the epidermal growth factor receptor
A:Reference number: A60143; MUID:85182650; PMID:2985580
A:Accession: A60143
A:Molecule type: protein
A:Residues: 740-744, 'X', 746-747 <RUS>
R:Motczkowski, B.; Mosig, G.; Cohen, S.
Nature 309, 270-273, 1984
A:Title: ATP-stimulated interaction between epidermal growth factor receptor and supercoiled DNA
A:Reference number: A38023; MUID:84191554; PMID:6325948
A:Contents: annotation; receptor activity

A:Molecule type: mRNA
 A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>
 A:Cross-references: EMBL:Z12608
 R:Heisermann, G.J.; Gill, G.N.
 J. Biol. Chem. 263, 13152-13158, 1988
 A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
 A:Reference number: A28941; MUID:88330814; PMID:3138233
 A:Accession: A28941
 A:Molecule type: protein
 A:Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000, 1002-1009,
 R:Hisbbs, M.L.; Dunn, A.R.; Alexander, W.S.
 submitted to the EMBL Data Library, April 1994
 A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
 A:Reference number: S45325
 A:Accession: S45325
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-971, 'K', 973-1210 <VER>
 A:Cross-references: EMBL:X78987; NID:G488830; PIDN:CAA55587.1; PID:G488831
 R:Paria, B.C.; Das, S.K.; Andrews, G.K.; Dev, S.K.
 Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
 A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse b
 A:Reference number: 149643; MUID:93126380; PMID:7678348
 A:Accession: 149643
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 12-20, 22-132 <RES>
 A:Cross-references: GB:L06864; NID:G193001; PIDN:AAA53029.1; PID:G567201
 C:Genetics:
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:648-670/Domain: transmembrane #status predicted <TM>
 F:712-977/Domain: protein kinase homology <KIN>
 F:720-728/Region: protein kinase ATP-binding motif
 F:680, 695/Binding site: phosphate (Thr) (covalent) #status experimental
 F:697, 1070, 1071/Binding site: phosphate (Ser) (covalent) #status experimental
 F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
 F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
 F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 45.4%; Score 3093; DB 2; Length 1210;
 Best Local Similarity 49.1%; Pred No. 1, 2e-119;
 Matches 625; Conservative 170; Mismatches 367; Indels 110; Gaps 23;

Qy 11 LLLALLPGAA--STQVCTGTDMLKRLPASPETHLDMLRLHLYQCGVQVQGNLEITYLPTN 68
 Db 14 LLTALCAAGGALEKKVCGQTSNRLTQLGTGFEDHFLSLQRMVNNCEVVLGNLEITYVQRN 73

Qy 69 ASLSFLQDIOEVOQYIKANSFEIGTELQRLRIVRGTOIFEDNYALAVLDNGDPLNNTTP 128
 Db 74 YDLSEFLKIQEAVGYVLLALNTVERIPLENLQIRGNALYENTVALATLSN----- 124

Qy 129 VTGASPGGLRLQLRLSLTEILKGGVLIQKNPOLCVQDITLWKDI----FHKNNOLALTLI 184
 Db 125 -YGNRTGLRLPMLNLEQLIGAVRPSNNPILCNMTDIQRDVRVQVFNMSMDL---- 180

Qy 185 DTRNSRACHPCSPMCKGRWGSESSDCQSLTRTVACGGCA-RCKGPLPTCCHEQCAAG 243
 Db 181 -QSHPSSCPCKDPSCPNGSCWGGGBCNQKLTKIICAQCCSHRCGRSPSCCHNQCAAG 239

Qy 244 CTGPKHSDCLAHNHSIGICELHCPALVTYNTDTFESMPNPEGYTFGASCVTACPNY 303
 Db 240 CTGPRESCLVCQKQFDEATKDTCPPLMLYNPTTYQMDVNPGEKYSFGATCVKCKPNY 299

Qy 304 LSTDVGSCTLVCLPLHNQVTAEDGTQRCCKSKPCARCYGLGMEHLRVAVTSANIQE 363
 Db 300 VYDHGSCVRACGPPYIEV-EDGIRKCKCKDGPCKVNGIGIGEPKDTLSINATIKH 358

Qy 364 FAGCKKIFGSLAFLPESFDGPASNTAPLOEQLOVFTLEITGYLVISAWPDSLPDLS 423
 Db 359.FKYCTAISGDLHLPLVAFKGSFTPTPLPRELEILKTVKEITGFLLIQAWPNWTDLH 418

RESULT 6

TVCHLV

epidermal growth factor receptor precursor - chicken

N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB

C;Species: Gallus gallus (chicken)

C;Date: 28-Feb-1996 #sequence_revision 05-May-1995 #text_change 04-Feb-2000

C;Accession: A27720; A00643

R;Lax, J.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennstr

Mol. Cell. Biol. 8, 1970-1978, 1988

A;Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou

A;Reference number: A27720; MUID:88261272; PMID:3260329

Qy 424 VFQNLQVIRGRILHNGAYSILTLQGLGTSWGLRSLRELGLSLALIHHTNLHLCFVHTVPWD 483
 Db 419 AFENLEIIRTKHQGFSLAVGLNITSLGLSLKEISDGVIIISGRNLVCYANTINWK 478

Qy 484 QLFERNPQALLHTANRPEDECVEGLACHOLCARGHCWGPGPTQCVCNCSQFLRGQECVEE 543
 Db 479 KLFGTNPQKTKIMNRAEKDCKAVNHVNCNPLCSSEGCWGPEPRDCVSCQVSRGRCVEK 538

Qy 544 CRVLQGLPRYVNAHCLPCHPCQPOQNGSVTCFGEPAEDOCVACHYKDPFCVACPSG 603
 Db 539 CNILEGEPRFVENSECICQHPCLPQAMNITCTGRGPDNCIOCAHYIDGPHCVKTCPCAG 598

Qy 604 VKPDLSTMPYTWKPEDEGACQPCINCTHSCVDLDDKGCPEAQORASPLTGISVAVVGILL 663
 Db 599 IMGENNTL-VKYNADANNVCHLCHANCYTCAGPGLOGCEVWPSPKPIPIATGIVUGLL 657

Qy 664 VVVLGVVFGI-LIKRRQOKIRKYTMRLLOTELVEPLTPSGAMPNQAOIRLKETELRK 722
 Db 658 FIVV-VALGIGLFWRRHIVKRTLRLLQERELVEPLTPSGEAPNQAHRLKETEFKK 716

Qy 723 VKVLGSGAFGVVKGWIIPGENVKIPVAIKVLRNTSPKANKIILDEAYVMAGVSGPYV 782
 Db 717 IKVLGSGAFGVVKGWIIPGEKVKIPVAIKELREATSPKANKIILDEAYVMASVDNPHV 776

Qy 783 SRLIGICLTSTVOLVTOLMPYGLLDHVRNRRGLSGQDLLNMCQIAKGMSYLEDVRLV 842
 Db 777 CRLLGICLTSTVOLVTOLMPYGLLDYVREHKONIGSYLLNCVQIAKGNVLEDRRLV 836

Qy 843 HRDLAARNVLKSPNPKVITDFGLARLLDIDETEHADGGKVPKIMMALESILRRRTHQ 902
 Db 837 HRDLAARNVLKTPQHVKITDFGLAKLGAEEKEYHAEGKVPKIMMALESILHRIVTHQ 896

Qy 903 SDVMSYGVTVWELMTFCAKYDGIIPAREIPDLLEKGERLPQPPICTTIDVYMIWKCMID 962
 Db 897 SDVMSYGVTVWELMTFGSKYDGIIPASDISILEKGERLPQPPICTTIDVYMIWKCMID 956

Qy 963 SECRPREFRLVSFRWARDPQFVVIQ-NEDLGPASPLDSTFYRSLLEDDMGDLVDAE 1021
 Db 957 ADSPKPRELILFSGWQARDPQVLYVQGERMHLPSPTNSFVRLMDEDMEDVVDAD 1016

Qy 1022 EYLVPPQGGFCPPDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEAAPRSLAPSEGA 1081
 Db 1017 EYLIPOQGGF-----NSPST-----SRTPLLSLSA 1042

Qy 1082 GSDVFDGDLGMAAKGLQSLPHTDPSLPQRYSDPTVPLPSET--DGYVAPLTCSPQPEY 1139
 Db 1043 TSN----NSTVACINRNGSCRVKEDAFQRYSSDPTGAVTEDNIDDAFL-----PVPEY 1092

Qy 1140 VNQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVWQVAFGGAIVENPEYL-T 1198
 Db 1093 VNQ-SVPRPAGSVQNPVYINQPLHP-----APGRDLHYQN--PHSNVAGNPEYLNT 1141

Qy 1199 PQGGAAPQHPHPAFSPAFDNLVYWDQ-----DP-----PERGAPPSFTKGTPT 1242
 Db 1142 AQ-----PTCLSSGFSNPAWIKGSHOWSLDNPDYQDDFFPKETKPGIFKG-PT 1191

Qy 1243 AENPEYLGLDVP 1254
 Db 1192 AENAEYLRVAPP 1203

A:Accession: A27720
 A:Molecule type: mRNA
 A:Residues: 1-1223 <LAX>
 A:Cross-references: GB:M20386
 R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
 Cell 41, 719-726, 1985
 A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
 A:Reference number: A00643; MUID:85228222; PMID:2988784
 A:Accession: A00643
 A:Molecule type: mRNA
 A:Residues: 585-1223 <NIL>
 A:Cross-references: GB:M10066
 C:Genetics:
 A:Gene: erbB
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
 C:pecific protein kinase
 F:1-30/Domain: signal sequence #status predicted <SIG>
 F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
 F:31-654/Domain: extracellular #status predicted <EXT>
 F:81-307/Domain: EGF receptor extracellular domain repeat <EE1>
 F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>
 F:655-677/Domain: transmembrane #status predicted <TM>
 F:678-1223/Domain: intracellular #status predicted <INT>
 F:719-984/Domain: protein kinase homology <KIN>
 F:727-735/Region: protein kinase ATP-binding motif
 F:136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #s
 F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
 F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
 F:754/Active site: Lys #status predicted
 F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 45.3%; Score 3085.5; DB 1; Length 1223;
 Best Local Similarity 48.1%; Pred. No. 2.5e-119;
 Matches 624; Conservative 177; Mismatches 351; Indels 145; Gaps 25;

Qy	8	RWGLLLALLPGAA-----STQCTGDMKRLRLPASPEHDLMLRHLQCVQVQGNLE	61
Db	13	RGAAVLVLLGLGVALCSAVEKKVCQGTNNKLTQLGHVEDHFTSLQRMNNCEVVLNLE	72
Qy	62	LTLYPTNASLFLQDIOEQVOYIKANSKFGITELQRLIRVRGTLQEDNVALAVLNGD	121
Db	73	ITYVEHRDITFLKTIQEVAGYVLIAMNMVDVIFLENLQIIRGNVLVNSFALAVLSNH	132
Qy	122	PLANNTPVTGASPGGLREQLRLSLEILKGGVILQRPOLCYQDTILWKDIFHKNOLAL	181
Db	133	-MNKIQ-----GLRLPKMLRLSEILNGVKLSNNPKLCMDTIVLNDIIDSRLK-PL	182
Qy	182	TLID-TNRSRACHPCSPMKSGRCWGESSEDCQLTRTVACGCA-RCKGPLPTDCCHEQ	239
Db	183	TVLDFASNLSCPKCHPNCETEDHCWAGAQNCQTLTKVICAQCSGRGRKVPDCCCHQ	242
Qy	240	CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAC	299
Db	243	CAAGCTGPRSDCLACRFRDADTKCTCPPLVLYNPPTYQMDVNPPEGKYSFGATCYREC	302
Qy	300	PYNLYSTDVSGCTLVCPHNEVTAEDGTORCEKSPCARVCYGLGMEHLREVRVATSA	359
Db	303	PHNYVTDHSGVRSVCNTDYEV-EENGVRCKKCDGLCKVCNGIGIGELKGLISINAT	361
Qy	360	NIQFAGCKKIFGLAFLPESFDGDPASNTAPLOEQLOVFETLEETIGVLIYISAWPDSL	419
Db	362	NIDSKNCTKINGDVSILPVAFLGDAFTKTLPLDPKLDVPRTVKEISGFLLIQAWFDNA	421
Qy	420	PDLVSFQNLQVIRGILHNGAYSITLQGLGISWLGSLRLSLGSLALIHNNHLCFVHT	479
Db	422	TDLYAFENLEIRTKHQGOYSLAVNLKIQSLGLSLKLEISDGDIAIMKNKNLCYADT	481
Qy	480	VPWQDLFRNPHOALLHTANRDEDCVGEGLACHOLCARGHCWGPQTCVNCOSFLRGQE	539
Db	482	MNWSRLFATOSQTKTIQNRRNKNDCTADRHVCDPLCSDVCGWCGPFPCHFSRFFSQKE	541
Qy	540	CVEECRVQLGLPREYVVARHCLPCHPECPQNG---SVTCFGRPADQCVACAHYKDPFFC	596

RESULT 7

A47253

epidermal growth factor receptor, HER4 - human

C:Species: Homo sapiens (man)

C>Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999

C:Accession: A47253

R:Flouman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; Ne

Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993

A:Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal

A:Reference number: A47253; MUID:93189574; PMID:838336

A:Accession: A47253

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-1308 <PLO>

A:Cross-references: GB:I07868; NID:G337359; PIDN:AAB59446.1; PID:G337360

A>Note: sequence extracted from NCBI backbone (NCBI:P126842)

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; growth factor receptor

F:716-981/Domain: protein kinase homology <KIN>

F:724-732/Region: protein kinase ATP-binding motif

Query Match	43.3%;	Score	2950.5;	DB 2;	Length	1308;	
Best Local Similarity	45.0%;	Pred. No.	8.9e-114;	Indels	173;	Gaps	28;
Matches	606;	Conservative	183;	Mismatches	384;		
QY	9	WGLLLALLPPGAA	---STQVCTGTGDKMLRLPASPTHLDMLRLHYLQGCQVVOGNLELTY	64			
DB	8	WVWSLLVAAGTVQPSDSQSVCAGTENKLSLSLDEQQYRALRKYYENCEVVMGNLEITS	67				
QY	65	LPTNASLSFLQIDIEVOOYIKANSKFIGITBLQRLRIVRGTHQFEDNVALAVLDNGDPLN	124				
DB	68	IEHRDLISFLRSREVITGVYLVALNQRYFLPLENLIIRGTKLIEDRYALAIFLNRYKDG	127				
QY	125	NTTPVTGASPGGLRELQRLSLEILKGVGLVLRNPOLCYQDTILWKDIFHKNNQLATLI	184				
DB	128	NF-----GLOELGLKUNITEILNGVYVDQNKFLCYADTIHWQDIVRNWPSNLTIV	178				
QY	185	DTRNSRACHPCSPMKSCRCRSGESSEDCQSILTRIVTCAGGC-ARCKGPIPTDCHEQCAAG	243				
DB	179	STNGSGGCRCHKSTG-RWCPTENHCQTLTRIVCAEQCDRCVGPVSDCHRCACAG	237				
QY	244	CTGPKHSCLACLHFNHSGICELHCPALVTVNTDTFESMPNDEGRYTFGASCVTACPNY	303				
DB	238	CSGPKDITDCFACMNFNSGACVTCQPTFFVYNPTTFQLEHNFNAKYTYGAFCKKCPHNF	297				
QY	304	LSTDVGSCTLVCLPHNQSVTAEDGTQCEKCKSPCARVCYGLGMEHLREVAVTSANIQE	363				
DB	298	V-VDSSSCVRAPSPSKMEV-EENGIMCKPCTDIPCACDGIPTGSLMSAQTVDDSSNIDK	355				
QY	364	FAGCKIFGSLAFAPESPDGPASNTAPLOEQVFELEIEITGILYISAWPDSLPDLS	423				
DB	356	FINTCKINGNLIFLVTHGHPYNAIEADPEKLVNFRTVEITGFLNIQSPNMTDFS	415				
QY	424	VFQNLQVIRGLHNGAVSLTQIGISWGLRSLRGLSGLALIHNNHLCFVHTVPWD	483				
DB	416	VFSNLVTIGRVLVSLGLSLLILKQGITSLQFQSLKEISAGNIYITDINSNLCTYHTINWT	475				
QY	484	QLFRNPHQALLHTANRPEDECVGEGLACHOLCARGCHGPGPTQVCNCSQFLRGOECVEE	543				
DB	476	TLFTSTINQRIVIRDRNKAENCTAEGWCVNHLCCSDGCGWPGPDQCLSCRRFSRGRICIES	535				
QY	544	CRVLQGLPREVYNARHCLFCHPECOF-QNGSVTCFGEADOCVCAHYKDPFPFCVARCP	602				
DB	536	CNLYDGEFERENGSIQVCEPDQCEKMEGDLTCHGPGDNTCKSHKFDGPNCKVCPD	595				
QY	603	GVPKPLSYNPIWKFDPDEGACQCPINCTHSCVDLDDKGC-----PACORASPL	651				
DB	596	GLOGANSE--IPKYADPDRECHPCPNCTQCGNGPTSHDCIYYPWTGHTSLPQHAR-TPL	652				
QY	652	TSIYSAVV-GILLVVVLGVVFGILIKRQOKIRKYTWERLLQETELVEPLTPSGAMPNOA	710				
DB	653	--IAAGVTGGULFVILVGLTFAVYVRKSIK-KKALRRFL-ETELVEPLTPSGTAPNOA	708				
QY	711	QMRILKTELKRVKVLGSAFTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE	770				
DB	709	QURLIKTELKRVKVLGSAFTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE	768				
QY	771	AYWAGVGSPPVSRLLGICLSTVOLVTOLMPYGLDLHVRNRLGSLQDLNWCQIA	830				
DB	769	ALIMASMDHPHLVRLGLVCLSPITQVLTQMLPHGCLLEYVHEHKDNIGSQLLLNWCQIA	828				
QY	831	KMSVLEEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEXHADGKVPKQMA	890				
DB	829	KGMVYLERRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEXHADGKVPKQMA	888				
QY	891	LESILRRFTTHQSDVMSYGVTVWELMTFCAPYDGPAREIPDLLEKGERLPQPPICTID	950				
DB	889	LECIVHFKTHQSDVMSYGVTVWELMTFGKPYDGPAREIPDLLEKGERLPQPPICTID	948				
QY	951	VTMIVMKWMDSECRPRELUSEFSRMARDPQRFVQINED-LGPAASPLDSTFYRSL	1009				
DB	949	VTMIVMKWMDIADSRPFKELAEAFSRMARDPQRYLVIQGDNRMLKLPSPNDSKFFQNL	1008				
QY	1010*	EDDDMGDLVDAEYLVPOQGFPCPDPAFCAGGVMVHRRSSSTRSGGDLTLGLPSEEE	1069				

Db	1009	DEEDLEDMDAEBEYLV-QAFNIPPP-----IYTSRARIDSNRS-----EIGHSPPPAY	1056				
Qy	1070	APRS-----PLAP-SEGAGSDVDFDGLMGAAKCLQS	1100				
Db	1057	TPMSGNQFVYVYRGGFAAEQGVSVYPRAPTSTIPEAPVAQGATAEIFFDDSCCNGLRKPVA	1116				
Qy	1101	LPTHDSPLQRYSEDPTVPLPS-----ETDGVAPLTCSPQPEYVQPDVAPQPPSPR	1153				
Db	1117	PHVQDSSTQRYSDPTVFAPERSPGELDEEGYMTFMRDKPQOEYLNPNVE-----	1167				
Qy	1154	EGPLPAARPAAGATLERAKTILSPGKNGQVDFVAFGGAVENPEYLTTPQGAAPQHPPPA-	1212				
Db	1168	ENFVSR-----KNGDLQ-----ALDNPEYHNASNG-----PKAE	1199				
Qy	1213	-----FSAFDNLYYWDQDPPPERGA--PPSTF	1237				
Db	1200	DEYVNEPLYLNTFANTLGKAEYLNKNNILSMPEKAKAFDNPDMYNNHSLPRSTLQHPDYL	1259				
Qy	1238	KGTPT-----AENPEYL	1249				
Db	1260	QYSTKYFYKQNGRIRPIVAENPEYL	1285				
RESULT 8							
S06142							
Protein-tyrosine kinase (EC 2.7.1.112) mrk-Y precursor - southern platyfish							
N;Alternate names: epidermal growth factor receptor homolog; kinase-related transforming							
C;Species: Xiphophorus maculatus (southern platyfish)							
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000							
C;Accession: S06142; S13809							
R;Wittbrodt, J.; Adam, D.; Mailitschek, B.; Maeueller, W.; Raulf, F.; Telling, A.; Roberts							
Nature 341, 415-421, 1989							
A;Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu loc							
A;Reference number: S06142; MUID:90015140; PMID:2797166							
A;Accession: S06142							
A;Molecule type: DNA							
A;Residues: 1-1166 <WIT>							
A;Cross-references: EMBL:X16891; NID:G65290; PIDN:CAA34770.1; PID:G65291							
R;Adam, D.; Maeueller, W.; Scharf, M.							
Oncogene 6, 73-80, 1991							
A;Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xiphophoru							
A;Reference number: S13807; MUID:91125882; PMID:1846957							
A;Accession: S13809							
A;Status: preliminary; translation not shown							
A;Molecule type: DNA							
A;Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>							
A;Cross-references: EMBL:X56319; NID:G65284; PIDN:CAA39763.1; PID:G65285							
C;Genetics:							
A;Gene: mrk							
A;Map position: Y							
A;Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1							
C;Superfamily: epidermal growth factor receptor; protein kinase homology							
C;Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein; tyro							
F;1-25/Domain: signal sequence #status predicted <SIG>							
F;26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>							
F;707-972/Domain: protein kinase homology <KIN>							
F;715-723/Region: protein kinase ATP-binding motif							
Query Match							
Best Local Similarity							
Matches							
39.0%; Score 2653; DB 1; Length 1166;							
44.8%; Pred. No. 1.2e-101;							
Matches 568; Conservative 163; Mismatches 398; Indels 138; Gaps 28;							
Qy	4	AALCRWGLLALLPPGAAT-----QVCTGTDMLKRLPASPTHLDMLRHLYQGCQVVOGN	59				
Db	8	AALLQ-----LLLVLSRCCSTDPDRKVCQGTSNQMT---LDNHYLKKMKMYSGCNVLEN	62				
Qy	60	LELYLPTNASLFLQIDIOEQVQVQYKANSFIGITLQRLRIVRGTHQFEDNVALAVLDN	119				
Db	63	LEITYTOENQDLSPLQSIQEVGGVLIAMNEVSTIPLNLRKQNLQYEGNFTLLVWSN	122				
Qy	120	GDPLNNTTPTVTGSPGGLRELQRLSLEILKGVGLVLRNPOLCYQDTILWKDIFHKNNQL	179				

Db 123 YQK-NPSSP--DVYQVGLKQLQLSNLTELISGGVVKVSHNPLLCNVETINWWDIVDKTSNP 179
Qy 180 ALTLIDNRSRACHPCSPMKGSCWGESSEDCSLTRTVCAAGC-ARCKGPLPTDCCHE 238
Db 180 TMNLIPIHAFERQCKQCHGCVNGSWAPGPGHCQKFTKLLCAEQCNRRRCRGPKPIDCNE 239
Qy 239 QCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNBEGRYTFGASCVTA 298
Db 240 HCAGCGTGPATDCCLACRDFNDGCTKDTCPPIKYDIVSHVQVVDNPNIKYTFGAACVKE 299
Qy 299 CPYNVLTDSGCTLVCPHMQEYTAEDGTORCEKSKPCARVCYGLGMEHLREVRAVTS 358
Db 300 CPSNVVTE-GACVRSACSAGMLEVD-ENGRKSCPKPCDGVCKVCDGIGISLNTIAVNS 357
Qy 359 ANIQEFAGCKKIFGLAFPLPESPDGPDASNTAPLQPEOLQVFETLEBITGYLISAWPDS 418
Db 358 TNIRSFNCTKINGDIILNRNSFEGDPHYKIGTMDPEHLNLTVTKEITGYLVIMWPFEN 417
Qy 419 LPDLISVQNLOVIRGRIHNGAYS-LTLQGLIGISWGLRSRELGLSLHHTHNLFCFV 477
Db 418 MTSLSVFQNLBIIRGRTTFSRFGSFVVQVVRHLQWGLRSLSKEYSAGNVILKNTLQLRYA 477
Qy 478 HTVPWDOLFNRPHOALLHTANRPEDECVGEGCLACHOLCARGHCWGPPTOCVNCQSQFLRG 537
Db 478 NTINWRLEFRSEDOSIEYDART-----ENQTCNNECSEDCGW-PGPTMCVSLCHVDRG 529
Qy 538 QECVEECRVLQGLPREYVNAHRLCHPECPQNGSVTCFGEADQCACAHYKDDPFVCV 597
Db 530 GRCVASCNLLQGEPREAQVDCRQVQCHQELVQDLSLTCYCGPGANCKSAHFQDGPQCI 589
Qy 598 ARCSGVKPDLSYMPIWKFPDEBACQPCPINCTHSCVDLDDKCPAEQASPLTSTVISA 657
Db 590 PRCPHGILGSDTL-IWKYADKMGCCQPCQNCQTCGSGPGLSGCRGD-IVSHSSLAVAL 647
Qy 658 VVGLLVVVLGVGILIKRQOKIRKYVTRILLQETELVEPLTPSGAMPNQAOIRLKE 717
Db 648 VSGLIITVIVALLVILRRRIK-RKRTIKCLQEKELVEPLTPSGQAPNQAFRLIKE 706
Qy 718 TELRKVKLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGV 777
Db 707 TEFKDRVLGSGAGFTYKGLWPDGENIRIPVAIKVLRNTPSKANKEILDEAYVMAGV 766
Qy 778 GSPVSVLLGICLTSTVOLVTQLMPYGLLDHVRNRRGLSQDILLNWCQIAKGMNYLE 837
Db 767 DHPHVCRLGICLTSAVLQVLTQLMPYGLLDYVRQHOERICQWLLNWCQIAKGMNYLE 826
Qy 838 DVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGCKYPIKWMALLESILRR 897
Db 827 ERHLVHRDLAARNVLLKNPNHVKITDFGLSKLLTADEKEYQADGCKYPIKWMALLESILQW 886
Qy 898 RFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVK 957
Db 887 TYTHQSDVMSYGVTVWELMTFGSKPYDGIIPAKETASVLENGERLPQPPICTIEVYIMLK 946
Qy 958 CWMIDSCRRFRRELVSFERSMARDPQRFVVIQNEIDLGPAASPLDSTFYRSLLEDGMDL 1017
Db 947 CWMIDPSRRFRRELVSFERSMARDPQRFVVIQNEIDLGPAASPLDSTFYRSLLEDGMDL 1001
Qy 1018 VDAEYLVPOGGFCPPDAPGAGGVVHRRSSSTRSGGDLTLGLFSPSEEAAPRSLAP 1077
Db 1002 VDAEYLLPYKRI-----NROGS-----EPCIP 1024
Qy 1078 SEGAGSDVFDGDLGMAAGKGLQLPHTDPSPLORYSEDPV-PLPSETDGVAPLTCSPQ 1136
Db 1025 PTGH-----PVRENSTLNIISDPTQNALEKLDGH----- 1055
Qy 1137 PEYVQPDVNRQP-----PSPRE-----GPLP-AARPAGATLERAKTISPGKNGVYKD 1183
Db 1056 -EYVNPQGETSSRLSDIYNPNVEDLTDGNGVPSLSQEAETNFSRPEYLNTNQNLSL--- 1111
Qy 1184 VFARGGAVENPEYLTPOGGAAPQHPAPSPAFDNLYYWDQDPERGAPSPFTKGTPTA 1243
Db 1112 PLVSSGSMDDPDY---QAG-----YQAAF-----LPOTGALTGMGLFPA 1149

Qy 1244 ENPEYLG 1250
Db 1150 ENLEYLG 1156

RESULT 9

A36223
C:Species: Homo sapiens (man)
C:Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C:Accession: A36223; I59164
R:Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A:Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal g
A:Reference number: A36223; MUID:90083234; PMID:2687875
A:Accession: A36223
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1342 <KRA>
A:Cross-references: GB:M29366
R:Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, G.J
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A:Title: Molecular cloning and expression of another epidermal growth factor receptor-re
A:Reference number: I59164; MUID:90311312; PMID:2164210
A:Accession: I59164
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>
A:Cross-references: GB:M34309; NID:G183990; PIDN:AAA35979.1; PID:G306841
C:Genetics:
A:Gene: GDB:ERBB3; HER3
A:Cross-references: GDB:119880; OMIM:190151
A:Map position: 12q13-12q13
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo
C:Keywords: ATP; phosphotransferase
F:707-972/Domain: protein kinase homology <KIN>
F:715-723/Region: protein kinase ATP-binding motif

Query Match 35.1%; Score 2389.5; DB 2; Length 1342;
Best Local Similarity 40.2%; Pred. No. 8.5e-91;
Matches 527; Conservative 189; Mismatches 466; Indels 129; Gaps 32;

Qy 10 GLLLALLPPGAA--STQVCTGTDMKRLPASPEHLMLRLHYQGGVOVGNLELTYLPT 67
Db 11 GLLFSLARGSEVGNQAVCPGTLNGLSVTGDENQVQTLKYKLCERCEVNGNLEIVLTGH 70
Qy 68 NASLSFLQDIQVQVYIKANSKFEIGITELQRLRIVRGTLFEDNYALAVLDNGDPLNNTT 127
Db 71 NADLSFLQWIREVTGYVLVANNEFTLPLNLRVVRGTQVYDGKFAIFVM----LYNNT 125
Qy 128 PVTGASPGGLRELQSLTELKGGVLIQBNPOLCYQDITLWKDIPKNNQLALTLDTN 187
Db 126 ----NSSHALRQURLTQLTEILSGGYVIEKNDKLCMDTIDWRD---AEIVVKD 178
Qy 188 RSRACHPCSPMKGSCWGESSEDCSLTRTVCAAGC-ARCKGPLPTDCHEOCAAGCTG 246
Db 179 NGRSCFPCEVCKG-RCWPGSEDCQTLTKTICAPQCNHGCFGNPNQCCHECAGCGSG 237
Qy 247 PKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNBEGRYTFGASCVTACPYNYLST 306
Db 238 PQDTCDFACRHFNDGACVPRCPQLVYNKLTLPQLEPNPHTKYQYGGVCVASCPhNFV-V 296
Qy 307 DVGSCITLVCPHMQEYTAEDGTORCEKSKPCARVCYGLGMEHLREVRAVTSANIOEPAG 366
Db 297 DOTSCVCRACPPDKMEVD-KNGLKMCPCGGLCPKACEGTGSG--SRFQTVSSNIDGFVN 353
Qy 367 CKKIFGSLAFPLPESPDGPDASNTAPLQPEOLQVFETLEBITGYLISAWPDSLPDLSVFQ 426
Db 354 CTKILGNLDFLITGLNGDPWHKIPALDPEKLVFRVREITGYLNTQSPHPHNFVSFVS 413
Qy 427 NLQVIRGRILHNGAYS-LTLQGLIGISWGLRSRELGLSLHHTHNLFCFVHTVPWDQL 485

Db 414 NLTTIGRSLYNRGSLIMKLNVTSLGRSLKEISAGRIYISANRQCLCYHSLNWKV 473
 Qy 486 FRNPHQALLHTA-NRPEDECVEGLACHOLCARGHCWPGTQCVCNCSQFLRGCEVCBEC 544
 Db 474 LRGPTEERLDIKHNRPRDCVAEGKVCDCPLSCSSGGCWGPGQCCLSCRNYSRGVGVCTHC 533
 Qy 545 RVLCGLPREYVYARHCLPCHPECPONGSVTCFPEADOCVACAHYKDPDPPCVCARCPGV 604
 Db 534 NFLNGEPREFAHEACSCHEPCQMGTCATNGSGSDTCAQACHFRDGPCHVCSSCHGV 593
 Qy 605 KPDLSYMPIWKFPEDEGACQPCINCHSCVDLDDKGCAPBQRA-----SPLTSTVSVAWG 660
 Db 594 LG--AKGSIYKYPDVQNECRPCHENCITQCGKGPQLQCLGQTLVLIGKTHLTALTVIAG 651
 Qy 661 ILLVVLGVVFGILIKRQOKIR-KYTHRRLLQETELVELPTPSGAMPNQAMRILKETE 719
 Db 652 --LVVIFMFLGGTFLYWRGRRIQNKRRMYRLERGESIEPLDPS--EKANKYLARIFKETE 708
 Qy 720 LRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKETLDEAYVMAGVGS 779
 Db 709 LRKVLGSGVFGTVHKGVWIPGESIKIPVCIKVIEDKSRQSFQAVTDHMLAIGSLDH 768
 Qy 780 PYVSRLLGICITSTVQLVTLQMPYGCCLLDHVRENRGLSGDOLLNWCMIQIAKMSYLEDV 839
 Db 769 AHIVRLGLCPGSSQLQVTLPLGLSLDHRQHRGALGPOLLNWNQVQIAKGMVYLEEH 828
 Qy 840 RLVRDLAARNVLKSPNHVKITDFGLARLIDIDETEHADGGKVPKIMMALESILRRFP 899
 Db 829 GMVHRNLAARNVLKSPSQVQVADFGVADLLPPDDKQLLYSEAKTPIKMALESILHFGKY 888
 Qy 900 THQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVWKW 959
 Db 889 THQSDVMSYGVTVWELMTFGAEPTAGLAEVPLLEKGERLAQPOICTIDVYIMVWKW 948
 Qy 960 MIDSECRPRFELVSEFSEMRARDPQFVVIQNEGLGPA---SPLDSTFYRSLLEDDMDGD 1016
 Db 949 MIDENIRPTEKELANEFTRMARDPPRYLVIKRES--GPGIAPGPSPHGLTNKKLEVELEP 1007
 Qy 1017 LVDAEYVLVPOQGFPCPPAPGAGGVVHRRSSSTRSGGDLTLGLEP--SEEAAPSPL 1075
 Db 1008 ELDDLDDLEAED-----NLATTLGSLALPVTGLNRPGRSOSLL 1048
 Qy 1076 APSEGAGSDVFDGLGMGAAGLQSLPTH--PSPLQRYSEDPTVPLP-----SETDGYV 1128
 Db 1049 SPSSGY-WPMNQNLGESCQSAVSGSSERCPVSLH-----PMPRGCLASSESEGV 1101
 Qy 1129 A-----PLTCSPOPE-----YVNPQDVPRPQPPSPREGP-----L 1157
 Db 1102 TGSEAEQLQKVMCRSRSRSPRPGDSAYHSQRHSLLTPTVPLSPGLSEEDVNGYVM 1161
 Qy 1158 PAARPAGATLERAKTLSP-GKNGV-----KDVPAFCGAVENPEYLTPOGGAAPQPHPP 1210
 Db 1162 PDTHLKGTSPSSREGTLSSVGLSSVLTGTEED-----EVEYMNRRRHSFP-PHPP 1212
 Qy 1211 PAFSPAFDNLVYWD-----ODPPERGAPPSTFKGTPTAENPEYL 1249
 Db 1213 RPSSLEELGEYEMDVGSLSASLSTQSCPLHPVIMTAGTTPDEYEM 1263

RESULT 10

JC4387
 epidermal growth factor receptor homolog precursor - rat
 N:Alternate names: ErbB3 protein; HER3 protein
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
 C:Accession: JC4387
 R:Hellyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
 Gene 165, 279-284, 1995
 A>Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.
 A:Reference number: JC4387; MUID:96096535; PMID:8522190
 A:Accession: JC4387
 A:Molecule type: mRNA
 A:Residues: 1-4339 <HEL>

A:Cross-references: GB:U29339; NID:G915389; PID:G915390
 A:Experimental source: liver
 A>Note: The authors translated the codon AAC for residue 369 as Thr and GTT for residue 370 as Trp. This protein is a functional heregulin receptor that transduces signals to the C:Genetics:
 A:Gene: ErbB3
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
 C:Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
 F:640-659/Domain: transmembrane #status predicted <TM>
 F:705-970/Domain: protein kinase homology <KIN>
 F:713-721/Region: protein kinase ATP-binding motif
 F:939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr) (C

Query Match 34.0%; Score 2316.5; DB 2; Length 1339;

Best Local Similarity 40.4%; Pred. No. 8.3e-88;
 Matches 518; Conservative 170; Mismatches 440; Indels 155; Gaps 34;

Qy 3 LAALCRWGLLLALLPGAA---STQVCTGTDMLRLPASPTHLDMLRHLVYQGCQVVOGN 59
 Db 7 LQVLC-----FLLSLARGSEMGNSQAVCPGLTGLSVTGDADNQYQTLTKLYEKEVVMGN 62
 Qy 60 LELTYLPTNASLFDIOEQVQYIKANSKFIIGITELQRLIRVGTOLFDENVALAVLDN 119
 Db 63 LEIVLTGHNADLSFLQWIREVYAVLVAMNEFSLPLPRLRVVRVGTQVYDGKFAIFVM-- 120
 Qy 120 GDPINNTPTVGASPGGLRELOLRSLTEILKGGVLIOQNPOLCYQDTILMKDIFKKNQL 179
 Db 121 ---LNYNT----NSSHALRQLKFTLITLSSGGVYIEKDKLCHMDTIDRVIDRVR-- 170
 Qy 180 ALTLIDTNRSRACHPCSPCKSRGSESSDQSLTRTVCAAGC--ARCKGPLPTDCHE 238
 Db 171 GAELVKNNGANCPPEVCKG--RCWGPDPDQILTKTICAPQCNGRCFNPQCHD 229
 Qy 239 QCAAGCTGPHSDCLACLFHNSGICELHCPALVYNTDTFESMNPNEGTYTGAASCTA 298
 Db 230 ECAGGCGPQDTDFACRRFNDGACVPRCPPELVVKNLTFOLEPNPHTKYQGGVCVAS 289
 Qy 299 CPNYLSTDVGSCTVCLPHNQVTAEDGTQRCCKSPCARVCYGLGMEHLREAVTS 358
 Db 290 CPNHFV--VDQTFVCRACFPDKMEVD--KHGLKMCPECGCLCPKACEGTGSG--SRYTQVDS 345
 Qy 359 ANIQEPAGCKKIFGSLAFLPESPDGPASNTAPLOPQLOVFETLEITGYLISAWPDS 418
 Db 346 SNIDGFVNTKILGNLDFLITGLNVDPWHKI PALDPEKLVFRTVREITGYLNIQSPPH 405
 Qy 419 LPDLSVFQNLQVIRGRILHNGAYS--LTLOGLIGISWGLRSLRELGSGLAIHHNTHLCFV 477
 Db 406 MNEFVSFNLTTIGRSLYNRGFSLLIMKLNVTSLGFRSLKEISAGRVYISANQQLCYH 465
 Qy 478 HTVPWDOLFNRPHQALLHTA--NRPEDECVEGLACHOLCARGHCWPGTQCVCNCSQFLR 536
 Db 466 HSLNWTLLRGPSEERLDIKYDRPLGCELAEGKVCDCPLSCSSGGCWGPAQCLSCRYSR 525
 Qy 537 GQECVEECRVQLGPREYVYARHCLPCHPECPONGSVTCFPEADOCVACAHYKDPDPPC 596
 Db 526 EGVCTVHCNLFQGEPRFVHEAQCFSPCLPMEGSTYNGSSDGCARCAHFRDGPHC 585
 Qy 597 VARPSCGVKPDLSYMPIWKFPEDEGACQPCPINCHTSC--VDLDDKCPAPQASPLTSI 654
 Db 586 VNSCPFGILG--AKGPIYKYPDAQNECPCHENCTQCGNPGELQDCLGQAEVLMSKPHLV 643
 Qy 655 VSAVVGILLVVLGVVFGIILKROQKIR-KYTHRRLLQETELVELPTPSGAMPNQAMR 713
 Db 644 IAVTVG--LAVILMILGSGFLYWRGRRIQNKRRMYRLERGESIEPLDPS--EKANKYLAR 700
 Qy 714 ILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKETLDEAYV 773
 Db 701 IFKETELRKVLGSGVFGTVHKGIWIPGESIKIPVCIKVIEDKSRQSFQAVTDHMLA 760
 Qy 774 MAGVGSYVSRLLGICITSTVQLVTLQMPYGCCLLDHVRENRGLSGDOLLNWCMIQIAKM 833

```
Db 761 VGSLDHAIHVLLGLCGSSQLQVTOYLPGLSLDHDVHQHRETLPGLLLNWGVQIAKGM 820
Qy 834 SYLEDVRLVRLDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALES 893
Db 821 YYLESHVMVRDLALRNVMKSPSQVQVADFGVADLLPPDDKQLLHSEAKTPIKMALES 880
Qy 894 ILRRFRTHQSDVWVGYVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYM 953
Db 881 IHFGKYTHQSDVWVGYVTVWELMTFGAEPYAGLRLAETPDLLEKGERLAQPOICITIDVYM 940
Qy 954 IMVKCMWIDSCRRFRFRELVESEFARMARDPQRFVVIQWEDLGPASPLDSTFYRSLLEDDDD 1013
Db 941 VMVKCMWIDENIRTFELANEFTRMARDPPRYLVIKRAS-CPGTP--PAAPSVLTKE 997
Qy 1014 MGDVDAEYLVPOQGFPCPDAPGAGGMVHHRSSSTRSGGDLTLGLPEPSEE----- 1068
Db 998 L-----QEALEPEL-----DLDLLEAEEGLATS 1023
Qy 1069 -----EAPRSLPASEG-----AGSDVPDGLGMGAAGKGLQSLPETHD 1105
Db 1024 LGSALSPTGTLTRPGSQSLSPSSGYMPMNQSLGEACLDVSLVGGREGQFSRPSILH- 1082
Qy 1106 PSPQRYSEDPTVLPSETDGVY---APL-----TC-----SPOPE-----VYNQPDV 1145
Db 1083 PIPGR-----PASESSEGHVTGSEAELOEKVSVCRSRSRSPRGRSDAETHQSHS 1135
Qy 1146 RPQPPSPREGP-----LPAARPAGATLERAKTLP-SGKNGVY-----KDVFAF 1187
Db 1136 LLTPVTLSPGLLEEDGNGVYMPDTHLGRASSREGTSSVGLSSVLGTTEDEED----- 1191
Qy 1188 GGAIVENPEYLTPOGGAAPQHPHP 1210
Db 1192 ----EYEYMNKRKRGSP-PRPP 1209

RESULT 11
TVFVLV
N;Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C;Species: avian leukosis virus, ALV
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C;Accession: B00643; A00643
R;Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A;Title: C-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: B00643
A;Molecule type: mRNA
A;Residues: 1-698 <N1>
A;Cross-references: GB:M10066; GB:M13881; NID:G211749; PIDN:AAA48763.1; PID:G211750
A;Note: in Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gal
C;Comment: This protein is synthesized as a gag-env-erbB protein.
C;Genetics:
A;Gene: gag-env-erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p
F;1-6/Product: gag protein (fragment) #status predicted <ENV>
F;7-59/Product: env protein (fragment) #status predicted <GAG>
F;60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F;194-459/Domain: protein kinase homology <KIN>
F;202-210/Region: protein kinase ATP-binding motif
F;229/Active site: Lys #status predicted

Query Match 25.9%; Score 1766.5; DB 1; Length 698;
Best Local Similarity 52.2%; Pred. No. 1.5e-65;
Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18;

Qy 578 GPEADQCVACAHYKDPFCFVARCPGSKVPLDLSYMPIWKFPDEEGACQPCPINTCHSCVDL 637
Db 60 GP--DHCMKCAHFIDGPHCVKACPAAGVLGENDTL-VMKYADANAVCOLCHPNCTRGCKGP 116
Qy 638 DDKGCPEAQERASPLTSTVSANV--GILLVVLGVVFGVILIKRQOKIRKYWRRLLOTEL 696
```

```
Db 117 GLEGCP---NGSKTPSTAAGVWGGLCLVAVVGLIGLYLRRR-HIVKRTLRRLLOREL 172
Qy 697 VEPLTPSGAMPNOAMRILKETELRKVKVGLSGAGFYVYKGIWTPDGENVKIPVAIKVLR 756
Db 173 VEPLTPSGEAPNQAHRLILKETEFKVKVGLSGAGFYVYKGIWTPDGENVKIPVAIKELR 232
Qy 757 ENTSPKANKEILDEAYVMAGVSPYVSRLIGICLTSTVQLVQLMYPYGCCLLDHVRNRR 816
Db 233 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMYPYGCCLLDYIREHKN 292
Qy 817 LGSQDLNLWCQWAKGMSVLEDLVLRDLAARNLVKSPNHVKITDFGLARLLDIDETE 876
Db 293 IGSQDLNLWCQWAKGMSVLEDLVLRDLAARNLVKTPQHVKITDFGLAKLLGADEKE 352
Qy 877 YHADGGKVPKIMMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLE 936
Db 353 YHAEGGKVPKIMMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASISSVLE 412
Qy 937 KGERLPQPPICITIDVYMIMVKCMWIDSECPFRFRELVESEFARMARDPQRFVVIQ-NEDLG 995
Db 413 KGERLPQPPICITIDVYMIMVKCMWIDADSRPKFRELIAEFKWARDPPRYLVIOGDERMH 472
Qy 996 PASPLDSTFYRSLLEDDMGDLVDAEYLVPOQGFPCPDAPGAGGMVHHRSSSTRSG 1055
Db 473 LPSPDTSKFRVTLMEBEDMEDIVDAEYLVPHQGF-----NSPST--- 513
Qy 1056 GGDLTGLLEPSEEAAPRSP-----APSEGAGSDVFDGDLGMGAAGKGLQSLPETHDPSPLQ 1110
Db 514 -----SRTPLLSLSATSNNNSATNCID-----RNGQGHVREDSEFVQ 550
Qy 1111 RYSEDDPTVLPSET--DGVYVAPLTCSPQPEYVNVQPDVVRPQPPSPREGPLPAARAGATLE 1168
Db 551 RYSSDPTGNFLESIDDGFL-----PAPEYVNVQ--LMPKKPS----- 585
Qy 1169 RAKTLPSPKNGVVKDVF-----AFGAVENPEYLTPOGGAAPQHPHPAPF 1213
Db 586 ----TAMVQNGIYNNISLTAISKLPMSDRYQNSHSTAVDNPEYL-----NTNQSPLA 633
Qy 1214 SPAPDNLVYWDQ-----DPPE-----RCAPPSTFGTPTAENPEYVLGLDVP 1254
Db 634 KTVFESSPYWQSGNHQINLDNPDYQODFLPNETKPNGLLKVPAEENPEYLRVNAAP 689

RESULT 12
TVYUHV
C;Species: avian erythroblastosis virus
C;Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C;Accession: A00644; A38022
R;Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A;Title: The erbB gene of avian erythroblastosis virus is a member of the src gene family
A;Reference number: A00644; MUID:84026539; PMID:6313229
A;Accession: A00644
A;Molecule type: DNA
A;Residues: 1-604 <YAM>
A;Cross-references: GB:K01216; NID:G209676; PIDN:AAA42400.1; PID:G209678
R;Debuire, B.; Henry, C.; Benaisa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Martin, J.
Science 224, 1456-1459, 1984
A;Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type of
A;Reference number: A38022; MUID:84223957; PMID:6328658
A;Accession: A38022
A;Molecule type: DNA
A;Residues: 1-28,'W', 30-139,'F', 141-145,'V', 147-152 <DEB>
A;Cross-references: GB:K02006
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p
F;130-395/Domain: protein kinase homology <KIN>
F;138-146/Region: protein kinase ATP-binding motif
F;165/Active site: Lys #status predicted
```



```

Qy 697 VEPLTPSGAMPNOQMRLKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVL 756
Db 114 VEPLTPSGEAPNOAHLRLKETEFKVKVLGFGAGFTVYKGLWIPEGEKVTIPVAIKEL 173
Qy 757 ENTSPKANKEILDEAYVMAGVSPYVSRLLGICLTSTVQLVTQLMPYGCCLLDHVREN 816
Db 174 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCCLLDYIREHK 233
Qy 817 LGSQDLNWCQIAKMSYLEDDRLVHRDLAARNVLKSPNHVKITDPGLARLLDIDETE 876
Db 234 IGSQYLLNWCQIAKGMNYLEERHLVHRDLAARNVLKTPQDVKITDFGLAKQLGADEKE 293
Qy 877 YHAGGKVPKIMMALESILRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLE 936
Db 294 YHAGGKVPKIMMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIIPASEISSVLE 353
Qy 937 KGERLPQPPICTIDVYIMVWCWMIDSECRPRELSEFSESRMARDPQRFVVIQ-NEDLG 995
Db 354 KGERLPQPPICTIDVYIMVWCWMSDADSRPKRELIAEFSEKWARDPPRYLVIOGDERMH 413
Qy 996 PASPLDSTFYRSLLEDDMGDLVDAEYLVFQQGFFCPDPAPGAGGMVHHRSSSTRSG 1055
Db 414 LPSPTDSKFYRTLMEEDMEDIVDAEYLVPHQGFF-----NSPST--- 454
Qy 1056 GGDLTGLLEPSEEAAPRSL-----APSEGAGSDVFDGDLGMAKGLQSLPTHDPSPQL 1110
Db 455 -----SRTPLSSLSATSNNSATNCIDRNGG-----H----- 481
Qy 1111 RYSEDPTVPLPSETDGYVAPLTCSPQPEYVNPQDVVRPQPPSPREGPLPAARPAGAT-LER 1169
Db 482 -----PVREGFL-----PAPEYVQ--LMPKKPSTAMVQNOIYNYISITAIK 523
Qy 1170 AKTLPQKNGVVKDVAFGGAVENPEYL 1197
Db 524 LPMDSRYN-----SHSTAVDNPEYL 544

```

Search completed: July 22, 2003, 09:08:02
Job time : 31.9062 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 22.9062 Seconds
(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-103-117-12
Perfect score: 6809
Sequence: 1 MEALACRWGLLLALLPPGA.....TFKGTPTAENPEYLGDPVP 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73.*

1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6730	98.8	1255	1 A24571	protein-tyrosine k
2	5957	87.5	1260	1 TVRTWU	protein-tyrosine k
3	5941.5	87.3	1254	2 I48161	p-185 precursor -
4	3160	46.4	1210	1 GQHUF	epidermal growth f
5	3130	46.0	1210	2 A53183	epidermal growth f
6	3113.5	45.7	1223	1 TVCHLV	epidermal growth f
7	3003.5	44.1	1308	2 A47253	epidermal growth f
8	2699	39.6	1166	1 S06142	protein-tyrosine k
9	2421.5	35.6	1342	2 A36223	kinase-related tra
10	2336.5	34.3	1339	2 JC4387	epidermal growth f
11	1766.5	25.9	698	1 TVFVLV	protein-tyrosine k
12	1703	25.0	604	1 TVYUHV	protein-tyrosine k
13	1654.5	24.3	1330	1 GQFFE	epidermal growth f
14	1647	24.2	544	2 S35745	protein-tyrosine k
15	1640	24.1	545	2 S00727	kinase-related tra
16	1623	23.8	540	2 B44776	protein-tyrosine k
17	1621	23.8	540	1 TVFVBB	protein-tyrosine k
18	1521	22.3	644	2 A36325	epidermal growth f
19	1311	19.3	1323	2 E88257	protein let-23 [im
20	1311	19.3	1374	2 S70712	protein-tyrosine k
21	1229	18.0	1369	2 S70713	protein-tyrosine k
22	1165	17.1	1717	1 A45558	epidermal growth f
23	1145	16.8	527	2 A42032	epidermal growth f
24	999.5	14.7	843	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	protein-tyrosine k
26	754.5	11.1	311	2 S13808	protein-tyrosine k
27	733	10.8	1363	2 T43220	insulin-like growt
28	720	10.6	1382	1 INHUR	insulin receptor p
29	713	10.5	1383	2 A36080	insulin receptor p

ALIGNMENTS

RESULT 1

A24571

protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N;Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein erb

C;Species: Homo sapiens (man)

C;Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999

C;Accession: A24571; A25491; A44188; B44188; I59509; I57622

R;Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;

Nature 319, 230-234, 1986

A;Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth

A;Reference number: A24571; MUID:86118663; PMID:3003577

A;Accession: A24571

A;Molecule type: mRNA

A;Residues: 1-1255 <YAM>

A;Cross-references: GB:X03363; NID:g31197; PIDN:CAA27060.1; PID:g31198

R;Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.

Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A;Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epider

A;Reference number: A25491; MUID:86016729; PMID:2995967

A;Accession: A25491

A;Molecule type: DNA

A;Residues: 737-1031 <SEM>

A;Cross-references: GB:M11767; NID:g182163; PIDN:AAA35808.1; PID:g553282

R;Cousens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg,

Science 230, 1132-1139, 1985

A;Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromo

A;Reference number: A44188; MUID:86070181; PMID:2399974

A;Accession: A44188

A;Molecule type: DNA

A;Residues: 740-910 <COU1>

A;Cross-references: GB:M12036; NID:g183988; PIDN:AAA35978.1; PID:g183989

A;Accession: B44188

A;Molecule type: mRNA

A;Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>

A;Cross-references: GB:M11730; NID:g183986

R;King, C.R.; Kraus, M.H.; Aaronson, S.A.

Science 229, 974-976, 1985

A;Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A;Reference number: I59509; MUID:85272597; PMID:2922089

A;Accession: I59509

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 832-909 <REX>

A;Cross-references: GB:I29395; NID:g459807; PIDN:AAA35809.1; PID:g459808

R;Tal, W.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987

A;Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptiona

A;Reference number: I57622; MUID:87286898; PMID:3039351

A;Accession: I57622

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-191 <TAL>

F:658-680/Domain: transmembrane #status predicted <TM>
F:723-988/Region: protein kinase homology <KIN>
F:731-739/Region: protein kinase ATP-binding motif
F:711,191,263,535,576,634/Binding site: carbohydrate (Aen) (covalent) #status predicted
F:691/Binding site: phosphate (Thr) (covalent) #status predicted
F:758/Active site: Lys #status predicted
F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 87.5%; Score 5957; DB 1; Length 1260;
Best Local Similarity 87.4%; Pred. No. 3.1e-237;
Matches 1098; Conservative 51; Mismatches 106; Indels 2; Gaps 2;

Qy 1 MELAALCRWGLLLALLPPGAASSTQVCTDMKRLPASPETHDMLRHLVGGCQVVOGNL 60
Db 4 MELAAWCRWGLLLALLPPGIAAGTQVCTDMKRLPASPETHDMLRHLVGGCQVVOGNL 63
Qy 61 ELTYLPTNASLSFLQDIQEVGYVLIHAHQVQVPLQRLVIRGTQGLFEDNVALAVLDNG 120
Db 64 ELTVPANASLSFLQDIQEVGYVLIHAHQVQVPLQRLVIRGTQGLFEDNVALAVLDNR 123
Qy 121 DPLANNQYKANSKF-IGITELQLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQL 179
Db 124 DPQNVAASTPGRTPEGLRELQRLSLTEILKGGVLIQRNPOLCYQDMVLMKDFRKNQL 183
Qy 180 ALTLIDNRSRACHPCSPMKGSRGWSSSDCSQSLRTVTCAGGCARCKGPLPTDCCHEQ 239
Db 184 APVDIDNRSRACHPCSPMKGSRGWSSSDCSQSLRTVTCAGGCARCKGPLPTDCCHEQ 243
Qy 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAC 299
Db 244 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTTC 303
Qy 300 PYNLSTDVGSCTLVCPILHNOEVTAEQGTQRCCKSPCARVCYGLGMEHLREVRVTS 359
Db 304 PYNLSTEVGSCTLVCPENNOEVTAEQGTQRCCKSPCARVCYGLGMEHLRGARITSD 363
Qy 360 NIQEPAGCKTIFGSLAFIPESFDGDPASNTAPLOEQVETLEETGYLYISAWPDSL 419
Db 364 NVQEFDGCKTIFGSLAFIPESFDGDPSSGIAPLRPEQLQVETLEETGYLYISAWPDSL 423
Qy 420 PDLVSFQNLQVIRGILHNGAYSITLQGLISWGLRSLRELGSGLALIHNTLHCFVHT 479
Db 424 RDLVSFQNLRIIRGILHNGAYSITLQGLISWGLRSLRELGSGLALIHNTLHCFVHT 483
Qy 480 VPWDLFRNPQALLHTANREDE-CVGEGLHACQLCARGHCWGPGTQCVCNCSFLRQG 538
Db 484 VPWDLFRNPQALLHSGNREDELCVSSGLVCNLSLCAHGCWGPGTQCVCNCSFLRQG 543
Qy 539 ECVEECRVQLGPREYVNRHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFCVA 598
Db 544 ECVEECRVKGLPREYVSDKRLPCHPECPQNSSETCFGEADQCAACAHYKDSSCVA 603
Qy 599 RCPGSKVPDLSYMPKPPDEEGACQPCINCTHSCVDLDDKGCAPASPLTSIYSAV 658
Db 604 RCPGSKVPDLSYMPKPPDEEGICQPCINCTHSCVDLDRGCPASRASPVTIATV 663
Qy 659 VGILLVVLGVVFGILIKRRQKIRKYTMRLLOETELVEPLTPSGAMPNQAORILKET 718
Db 664 EGVLFLILVVLGVVFGILIKRRQKIRKYTMRLLOETELVEPLTPSGAMPNQAORILKET 723
Qy 719 ELRKVKVLGSGAGFTVYKGIWIPGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGV 778
Db 724 ELRKVKVLGSGAGFTVYKGIWIPGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGV 783
Qy 779 SPYVSRLLIGLCTSTVOLVTOLMPYGCILLDHVNRNRELGSQDILLNWCMTAKMSVLED 838
Db 784 SPYVSRLLIGLCTSTVOLVTOLMPYGCILLDHVNRNRELGSQDILLNWCMTAKMSVLED 843
Qy 839 VRLVHRDLAARNVLVKSPPNHVKITDFGLARLLDIDETEHADGKVPKWMALSIILRR 898
Db 844 VRLVHRDLAARNVLVKSPPNHVKITDFGLARLLDIDETEHADGKVPKWMALSIILRR 903
Qy 899 FTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICTIDVTYIMVVC 958

Db 904 FTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICTIDVTYIMVVC 963
Qy 959 WMTDSECRPRFRELVSFSSMARDPQFVVIQNEIDLGASPLDSTFVRSILLEDDMGDLV 1018
Db 964 WMTDSECRPRFRELVSFSSMARDPQFVVIQNEIDLGASPLDSTFVRSILLEDDMGDLV 1023
Qy 1019 DAEYLVPOQGFPCPDAPGAGGMVHRHSSSTRSGGDLTLGLEPSEEEAPRSLAPS 1078
Db 1024 DAEYLVPOQGFPCPDAPGAGGMVHRHSSSTRSGGDLTLGLEPSEEEAPRSLAPS 1083
Qy 1079 EGAGSDVFCGLGMAAGKLOSLPTDPPSLQRYSEDPVPLPSETGYVAPLTCSPQPE 1138
Db 1084 EGAGSDVFCGLGMAAGKLOSLPTDPPSLQRYSEDPVPLPSETGYVAPLTCSPQPE 1143
Qy 1139 YVNPQDVRFPQPPREGPLPAAPAGATLERAKTSLPGKNGVKDVFAGGAVENPEVLT 1198
Db 1144 YVNPQDVRFPQPPREGPLPAAPAGATLERAKTSLPGKNGVKDVFAGGAVENPEVLT 1203
Qy 1199 PQGGAAPQHPHPPAFSPAFNLYWDQDPPERGAPESTFKGTPTAENPEYLGDDVPV 1255
Db 1204 PREGTASPPHPSPAFSPAFNLYWDQDPPERGAPESTFKGTPTAENPEYLGDDVPV 1260

RESULT 3
I48161
p-185 precursor - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C:Accession: I48161
R:Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Yamazaki, Y.; Ishikawa
Gene 140, 251-255, 1994
A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A:Reference number: I48161; MUID:94193007; PMID:7908275
A:Accession: I48161
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: mRNA
A:Residues: 1-1254 <RES>
A:Cross-references: GB:D16295; NID:G493236; PIDN:BAA03801.1; PID:G747595
C:Genetics:
A:Gene: neu
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif

Query Match 87.3%; Score 5941.5; DB 2; Length 1254;
Best Local Similarity 87.1%; Pred. No. 1.3e-236;
Matches 1093; Conservative 58; Mismatches 103; Indels 1; Gaps 1;

Qy 1 MELAALCRWGLLLALLPPGAASSTQVCTDMKRLPASPETHDMLRHLVGGCQVVOGNL 60
Db 1 MELAAWCRWGLLLALLSPGASGTQVCTDMKRLPASPETHDMLRHLVGGCQVVOGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVGYVLIHAHQVQVPLQRLVIRGTQGLFEDNVALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVGYVLIHAHQVQVPLQRLVIRGTQGLFEDNVALAVLDNR 120
Qy 121 DPLANNQYKANSKF-IGITELQLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQL 180
Db 121 DPLANNQYKANSKF-IGITELQLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQL 180
Qy 181 LTLDITNRSRACHPCSPMKGSRGWSSSDCSQSLRTVTCAGGCARCKGPLPTDCCHEQ 240
Db 181 PVDIDNRSRACHPCSPMKGSRGWSSSDCSQSLRTVTCAGGCARCKGPLPTDCCHEQ 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAC 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTTC 300
Qy 301 YNVLSTDVGSCTLVCPILHNOEVTAEQGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNVLSTEVGSCTLVCPILHNOEVTAEQGTQRCCKSPCARVCYGLGMEHLRGARITSAN 360

QY 361 IQEPAGCKKIRGSLAFIPESFDGPPASNTAPLQEQLOVFTLEBEITGYLYISAWPDSLP 420
Db 361 IQEPAGCKKIRGSLAFIPESFDGPPASNTAPLQEQLOVFTLEBEITGYLYISAWPDSLP 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLSLTQGLIGISWGLGRLSRLGSLGALIHNTLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLSLTQGLIGISWGLGRLSRLGSLGALIHNTLHLCFVHTV 480
QY 481 PWDQLFRPHQALLHTANRPDEBCVGBGLACHQICARHCCHGCPGPTQCVNCSQFLRGQEC 540
Db 481 PWDQLFRPHQALLHTANRPDEBCVGBGLACHQICARHCCHGCPGPTQCVNCSQFLRGQEC 540
QY 541 VEECRVLGGLPREYVYARHCLPCHPECOQNGSVTCFGEADQCVACAHYKDPFCVARC 600
Db 541 VEECRVLGGLPREYVYARHCLPCHPECOQNGSVTCFGEADQCVACAHYKDPFCVARC 600
QY 601 PSQVKPDLSPYPIWKFPEDEGACQPCINCTHSCVDLDKCPAEQASPLTSIVAVVG 660
Db 601 PSQVKPDLSPYPIWKFPEDEGACQPCINCTHSCVDLDKCPAEQASPLTSIVAVVG 660
QY 661 ILLVVVLGVFGIILKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVVVLGVFGIILKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
QY 721 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGP 780
Db 721 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGP 780
QY 781 YVSRLLGICLSTVOLVTQMLPYGCLLDHVRNRRGLSGQDLNMCQIAKMSYLEDDR 840
Db 781 YVSRLLGICLSTVOLVTQMLPYGCLLDHVRNRRGLSGQDLNMCQIAKMSYLEDDR 840
QY 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGCKVPIKMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGCKVPIKMALESILRRFT 900
QY 901 HQSDVWSGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPPICITDVTYMTVMKWM 960
Db 901 HQSDVWSGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPPICITDVTYMTVMKWM 960
QY 961 IDSECRPRFELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDGDLVDA 1020
Db 961 IDSECRPRFELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDGDLVDA 1020
QY 1021 EYLVPQOGFFCPDPAPGAGVHRRSSSTRSGGDLTLGLPSEEEARPLAPSEG 1080
Db 1021 EYLVPQOGFFCPDPAPGAGVHRRSSSTRSGGDLTLGLPSEEEARPLAPSEG 1080
QY 1081 AGSDVFDGDLGMAKGLQSLPTHDPSPLOQYSEDPTVPLPSETDGVYAPLTCSPQPEYV 1140
Db 1081 AGSDVFEGLNGATKGQSI SPRDLSPLQYSEDPTVPLPSETDGVYAPLTCSPQPEYV 1140
QY 1141 NOPDVVRPQPPSPREGPLPAARPAAGATLERAKTLSPKNGVVKDVPFAGGAVENPEYLTTPQ 1200
Db 1141 NOPEVRPQPPSPREGPLPAARPAAGATLERAKTLSPKNGVVKDVPFAGGAVENPEYLTTPQ 1200
QY 1201 CGAARPPHPPAFSPAFNLVYWDODPPERGAPSTFKGTATNPEYLGIDVVP 1255
Db 1201 GGSASQPH-PPALCAFDNLVYWDODPPERGAPSTFKGTATNPEYLGIDVVP 1254

RESULT 4

GQHU

N:Contains: growth factor receptor precursor - human

C:Species: Homo sapiens (man)

C:Date: 15-Nov-1984 #sequence revision 27-Nov-1985 #text change 11-Jun-1999

C:Accession: A00641; A25772; S30024; A38672; A00642; A43515; A23062; A05281; A60143; A33

R:Ulrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.; Y

rg, P.H.

Nature 309, 418-425, 1984

A:Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression of

A:Reference number: A00641; MUID:84219729; PMID:6328312
A:Accession: A00641
A:Molecule type: mRNA
A:Residues: 1-1210 <LL>
A:Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA25240.1; PID:g757924
A:Note: The authors translated the codon AAG for residue 540 as Asn
R:Ishii, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
A:Title: Characterization and sequence of the promoter region of the human epidermal gr
A:Reference number: A25772; MUID:85270438; PMID:2991899
A:Accession: A25772
A>Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-29 <ISH>
A:Cross-references: GB:M11234; NID:g181981; PIDN:AAA52370.1; PID:g553272
R:Haley, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ullrich, A.; Waterfield, M.
Oncogene Res. 1, 375-396, 1987
A:Title: The human EGF receptor gene: structure of the 110 kb locus and identification
A:Reference number: S30024; MUID:88217333; PMID:3329716
A:Accession: S30024
A:Molecule type: DNA
A:Residues: 1-29 <HA2>
A:Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
R:Haley, J.D.; Waterfield, M.D.
J. Biol. Chem. 266, 1746-1753, 1991
A:Title: Contributory effects of de Novo transcription and premature transcript termina
A:Reference number: A38672; MUID:91107677; PMID:1988448
A:Accession: A38672
A:Molecule type: DNA
A:Residues: 1-29 <HAL>
A:Cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271
A:Experimental source: carcinoma cell line A431-7
R:Xu, Y.; Ishii, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.; Me
Nature 309, 806-810, 1984
A:Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RNAs
A:Reference number: A00642; MUID:84245835; PMID:6330563
A:Accession: A00642
A:Molecule type: mRNA
A:Residues: 150-187 'KSVIOAV', 195, 'M', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-322
'798-799, 'TD', 802-811, 'R', 813-942 <XY>
A:Experimental source: A431 human carcinoma cells, which have large numbers of EGF rece
R:Lin, C.R.; Chen, W.S.; Krueger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verma, I.
Science 224, 843-848, 1984
A:Title: Expression cloning of human EGF receptor complementary DNA: gene amplification
A:Reference number: A43615; MUID:84196372; PMID:6326261
A:Accession: A43615
A:Molecule type: mRNA
A:Residues: 713-964 <LIN>
A:Experimental source: epidermoid carcinoma cell line A431
R:Simmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.
Biochem. Biophys. Res. Commun. 124, 125-132, 1984
A:Reference number: A23062; MUID:85046483; PMID:6093780
A:Accession: A23062
A:Molecule type: mRNA
A:Residues: 1028-1210 <SIM>
R:Weber, W.; Gill, G.N.; Speiss, J.
Science 224, 294-297, 1984
A:Reference number: A05281; MUID:84172183; PMID:6324343
A:Accession: A05281
A:Molecule type: protein
A:Residues: 25-30, 'S', 32-51, 454-467 <WEB>
R:Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.
J. Biol. Chem. 260, 5205-5208, 1985
A:Title: Identification of residues in the nucleotide binding site of the epidermal gr
A:Reference number: A60143; MUID:85182650; PMID:2985580
A:Accession: A60143
A:Molecule type: protein
A:Residues: 740-744, 'X', 746-747 <RUS>
R:Mroczkowski, B.; Mosig, G.; Cohen, S.
Nature 309, 270-273, 1984
A:Title: ATP-stimulated interaction between epidermal growth factor receptor and superc
A:Reference number: A38023; MUID:84191554; PMID:6325948
A:Contents: annotation; receptor activity


```
Db 123 -----YQKNPSPEDVYQVGLKQLQLSNLTLLSGGVKVSHPNLLCNVETINWMDIVDK 175
Qy 176 NNQALTLIDNRRACHPCSPMCKSGSCWGESSEDCQSLTRTVCAAGC-ARCKGPLPTD 234
Db 176 TSNPTMNLIPHAERQCKQKDHGCVNGSCWAPGPGHCQKFTKLLCAEQCNERCRGPKPID 235
Qy 235 CCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYTGAS 294
Db 236 CCNEHCAGGCTGPRATDCLACRDNFDDGTCRDKTCTPPPKIYDIVSHQVVDNPNIKYTGAA 295
Qy 295 CVTACPVNLTSDVGSCTLVCPHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLEVR 354
Db 296 CVKECPNYYVTE-GACVRSACSAGMLEVD-ENGRKSKPCDGVCPKCGDIGIGLSNTI 353
Qy 355 AVTSANTQEFAGCKKIFGSLAFLESFDDPASNTAPLQPPQLQVFTLEITGYLXISA 414
Db 354 AVNSTINRSPNCTKINGDIILNRNSFEGDPHYKIGTMDPEHLNLTIVKEITGYLVIMW 413
Qy 415 WPSLPLDSVQNLOVIRGRILHNGAYS-LTQGLIGISWLGRLSRLGSLALIHNTH 473
Db 414 WPNMTSLSVQNLEIRGTRTTFSGRGSFVVVQVRHLQWLGLRSLKEYSAGNVILKNTLQ 473
Qy 474 LCFVHTVPWDLFRNPHQALLHTANRDECVGSLACHQLCARGHGWGPGTQCVNCSQ 533
Db 474 LRYANTINWRLLFRSDQSIYDART-----ENQTCNCSBEGCW-PGPTMCVCSLH 525
Qy 534 FLRQECVEECRLVQLPREYVNAHCLPCHPEQCPQNGSVTCFPGPADQCVACAHYKDP 593
Db 526 VDRGRCVASCNLLQGEPREAQVDCRQVCHQELVQTDLSITCYGPGPANCSSAHPQDG 585
Qy 594 PFCVARCPGKVPDLSYMPIWKFDPDEGACQPCPINCTHSCVDDDDKGCAPAEQASPLTS 653
Db 586 PQCIPTCPHGTGLBGDYL-INKYADKMGCCQPCQHCNTQCGSGPGLSGCRGD-IVSHSSL 643
Qy 654 IVSAVVGILLVVLGVVFGILLKEROOKIRKYTWRRLLQETELVEPLTPSGAMPNQOMR 713
Db 644 AVGLVSGLLIITVALLIVLWLRRIK-RKRTIRCLLQEKELVEPLTPSGAQNQAPLR 702
Qy 714 ILKETELRKVVLGSGAGFTVYKGIWIPDGNVNKIPVAIKVLRNTPSPKANKEITLDAVY 773
Db 703 ILKETEFKORVLGSGAGFTVYKGLNPDGENIRIPVAIKVLRNTPSPKANKEITLDAVY 762
Qy 774 MAGVSPVSRLLGLCTSTVOLVTQLMPYQCLLDHVRNKRGLSQDLDLWCMQIAKGM 833
Db 763 MASVDHPHVCRLGLCTLSAQLVTQLMPYQCLLDYVRQHOERICGQWLLNWCQIAKGM 822
Qy 834 SYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMAL 893
Db 823 NYLBERHLVHRDLAARNVLLKNPNHVKITDFGLSKLLTADKEYOAGDGGKVPKWMAL 882
Qy 894 ILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM 953
Db 883 ILQWYTHQSDVMSYGVTVWELMTFGSKPYDGIIPAKEIASVLENGERLPQPICTIEVYM 942
Qy 954 IMVKWIMIDSECRPRFRELSEFMRMARDPQRFVVIQNEIDLGAPSLDSTFYRSLLEDD 1013
Db 943 IILKWMIDPSRRPRFRELVEGFQMARPSRYLVIQ-----NLPSSDRRLFSRLSSDD 999
Qy 1014 MGLDVAEEYLVPOOGFPCPDPAAGAGGVVHRRSSSTRSGGDLTLGLPSEEPAPRS 1073
Db 1000 --DVVDADAYLLPYKRI-----NRQGS-----E 1020
Qy 1074 PLAPSEAGSDVDFGDLGMAAGKGLQSLPHDPSPLOQYSBDPTV-PLPSETDGYVAPLT 1132
Db 1021 PCIPFTGH-----PVRENISITLNIISDPTQNALEKDLDDH----- 1055
Qy 1133 CSPOEYVNPQDVRPQP-----PSPRE-----GPLP-AARPAGATLERAKTLPQKNG 1179
Db 1056 -----EYVNPQGETSSRLSDIYNPNVEDLTDGMPVSLSSQEAETFSRPEYLTNQNS 1110
Qy 1180 VKDVFAGGAVENPEYLTPOGGAAPQHPPPAPSPAFDNLNYWDQDPPERGAPSTFKG 1239
Db 1111 L---PLVSSGSMDDPDY---QAG-----YQAAF-----LPQTGALTNGMF 1145
```

Qy 1240 TPTAENPEYLG 1250

Db 1146 LPAENLEYLG 1156

RESULT 9

A36223

kinase-related transforming protein (erbB3) (EC 2.7.1.-) precursor - human

C:Species: Homo sapiens (man)

C:Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000

C:Accession: A36223; I59164

R:Kraus, M.H.; Iseling, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.

Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989

A:Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal

A:Reference number: A36223; MUID:90083234; PMID:2687875

A:Accession: A36223

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1342 <KRA>

A:Cross-references: GB:M29366

R:Plowman, G.D.; Whittney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, G.

Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990

A:Title: Molecular cloning and expression of another epidermal growth factor receptor-r

A:Reference number: I59164; MUID:90311312; PMID:2164210

A:Accession: I59164

A:Status: preliminary; translated from GB/EMBL/DBDJB

A:Molecule type: mRNA

A:Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>

A:Cross-references: GB:M34309; NID:G183990; PID:AAA35979.1; PID:G306841

C:Genetics:

A:Gene: GDB:ERBB3; HER3

A:Cross-references: GDB:119880; OMIM:190151

A:Map position: 12q13-12q13

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homol

C:Keywords: Atp; phosphotransferase

F:707-972/Domain: protein kinase homology <KIN>

F:715-723/Region: protein kinase ATP-binding motif

Query Match

Best Local Similarity 35.6%; Score 2421.5; DB 2; Length 1342;

Matches 530; Conservative 192; Mismatches 460; Indels 129; Gaps 32;

Qy 10 GLLLALPPGAA--STQVCTGTDMLRPLASPETHLDMLRHLVGGCQVQGNLELTPLT 67

Db 11 GLFLSARGSEVGNSSQAVCPGTLNGLSVTGDENQVQTLKLYERCEVVMGNLEIVLTGH 70

Qy 68 NASLSFLDIOEQVGVYLIHNNQVRQVPLQRLRIVRGTLQFEDNYALAVLDNGDPLNQY 127

Db 71 NADLSFLQWIREVTGYVLVAMNBFSTLPLNLRVVRGTQVYDGKFAIFVMLN----- 122

Qy 128 IKANSKFIGITELQLRSLTEILKGGVLIQNPOLCQVDTTLWKDIFHKNQLALTLDTN 187

Db 123 YNTNSSH-AURQLRLTQLTEILSGGVYIEKDKLCHMDTIDWRDIDVRDRD---AEIVVKD 178

Qy 188 RSRACHPCSPMCKSGSCWGESSEDCQSLTRTVCAAGC-ARCKGPLPTDCCHCOAAGTG 246

Db 179 NGRSCPPCHVCKG-RCWGPSEDCQTLTKTIAPQCNGHCFGPNPNQCCHDECAGCGSG 237

Qy 247 PKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYTGASCTVACPYNLST 306

Db 238 PQDITDCFACRHFNDGACVPCRPQPLVYNKLTFLQEPNPHTKYQYGVGVASCPHNFV-V 296

Qy 307 DVGSCTLVCPHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLEVRVATYSANIQFAG 366

Db 297 DOTSCVRACPPDKMEVD-KNGLKMCBPCGGLCPKACEGTGSG--SRFTYDSSNIDGFVN 353

Qy 367 CKKIFGSLAFLESFDDPASNTAPLQPEQLQVFTLEITGYLYTSAMPDSLPLDSVFO 426

Db 354 CTKILGNLDFLITGLNGDPWHKIPALDPEKLVNFRVREITGYLTNTQSPHMHNSVFS 413

Qy 427 NLOVIRGRILHNGAYS-LTQGLIGISWLGRLSRLGSLALIHNTHLFCFVHTVPMDQL 485

Db 761 VGSIDHAHIVRLGLCPGSSQLQVLYPLGLSLLDHVKQHRETLGPQLLLNNGVQIAKGM 820
Qy 834 SYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGKVPKIMWALES 893
Db 821 YYLEEHSVHRDLALRNVLKSPSQVQVADFGVADLLPPDDKQLLHSEAKTPIKWALES 880
Qy 894 ILRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPQPICTIDVYM 953
Db 881 IHFGKYTHQSDVWSYGVTVWELMTFGAEPYAGLRLAEIPDLLEKGERLAQPICTIDVYM 940
Qy 954 IMVKWMIDSRCRRFRELVSERFMRMARDPQREYVIONEDIGPASPLDSTFYRSLLEDDDD 1013
Db 941 VMVKWMIDENIRTFKELANEFTRMARDPPRYLVIKRAAS-GPQTP--PAAEPSVLTKE 997
Qy 1014 MGDLDVAEYLVPOQGFPCPDPAFGAGGMVHRHRSSTRSGGDLTLGLPESBE----- 1068
Db 998 L-----QEALEPEL-----DLDDLLEABEEGLATS 1023
Qy 1069 -----EAPRSLPARG-----AGSDVFDGLGMAAKGLOSLPTH 1105
Db 1024 LGSALSUPTGTLTRPGSQSLSPSSGYMPMNQSSLGEACLDLSAVLGGREGQFSRPSILH- 1082
Qy 1106 PSPIQRYSDDPTVLPSETDGYV---APL-----TC-----SPOPE---VYNOPDV 1145
Db 1083 PIPEGR-----PASRSSEGHVTCSEAELOEKVSVCRSRSRSPRGRGDSAVHSQRHS 1135
Qy 1146 RPQPPSPREG-----LPAARPAGATLERAKTLSP-GKNGVV-----KVVFAP 1187
Db 1136 LLTPVTLSPGLEEDGNGVYMPDTHLRGASSREGTLSSVGLSSVLGTEEDED----- 1191
Qy 1188 GGAIVENPEYLTPOGGAAPQHP 1210
Db 1192 ---EYEVYMNKRKRGSP-PRPP 1209

RESULT 11
TVFVNV
N;Contains: tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
C;Species: avian leukosis virus, ALV
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C;Accession: B00643; A00643
R;Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A;Title: C-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro-
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: B00643
A;Molecule type: mRNA
A;Residues: 1-698 <Nil>
A;Cross-references: GB:M10066; GB:M13881; NID:G211749; PIDN:AAA48763.1; PID:G211750
A;Note: in Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gal
C;Comment: This protein is synthesized as a gag-env-erbB protein.
C;Genetics:
A;Gene: gag-env-erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p
F;1-6/Product: gag protein (fragment) #status predicted <GAG>
F;7-59/Product: env protein (fragment) #status predicted <ENV>
F;60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F;194-459/Domain: protein kinase homology <KIN>
F;202-210/Region: protein kinase ATP-binding motif
F;229/Active site: Lys #status predicted

Query Match 25.9%; Score 1766.5; DB 1; Length 698;
Best Local Similarity 25.2%; Pred. No. 1.2e-65;
Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18;

Qy 578 GPEADQCACAHYKDPFPCVACRPSGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDL 637
Db 60 GP--DHCWKAHFDGPHCVKACPAVLGENDTL-VMKYADANAVCQLCHPNCITRGCKGP 116
Qy 638 DDKCPAEQASPLTISVAVV-GILLVVLGVVFGILIKRQKIRKYMRRLQLQTEL 696

Db 117 GLSECP---NGSKTPSTAAGVVGGLCLVWVGIGIGLYLRRR-HIVKRTLRLLRLQREL 172
Qy 697 VEPLTPSGAMPNOAQRILKETELRKVKVLGSGAFGVYKGIWIPDCENKVPVIAIKVL 756
Db 173 VEPLTPSGEAPNOAHLRIKETEFKKVKVLGSGAFGVYKGLWIPGEKVKVIAIKEL 232
Qy 757 ENTSPKANKILDEAYVMAGVGSPPYVSRLLGICLTSTVQLVQLMPYGCLLDHVRENRR 816
Db 233 EATSPKANKILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDIYREHKDN 292
Qy 817 LGSODLLNWCWOTAKGMSYLEDLVRLHRLDAARNVLKSPNHVKITDFGLARLLDIDETE 876
Db 293 IGSOYLLNWCVCWOTAKGNNVLEERRLHRLDAARNVLKTPQHKVITDFGLAKLGLADEKE 352
Qy 877 YHADGGKVPKIMWALESILRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLE 936
Db 353 YHAEGGKVPKIMWALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIIPASEISSVLE 412
Qy 937 KGBRLPOPICTIDVIMVWKWMIDSECRPRFRELVSERFMRMARDPQRFVVIQ-NEDLG 995
Db 413 KGBRLPOPICTIDVIMVWKWMIDADSRPKFRELIAEFKMDARDPPRYLVITQGDERRMH 472
Qy 996 PASPLDSTFYRSLLEDDDDMGDLVAEYLVPOQGFPCPDPAFGAGGMVHRHRSSTRSG 1055
Db 473 LPSTTSKFRITLWEEDEMDIVDADEYLVPHQGF-----NSPST--- 513
Qy 1056 GGDLTGLPESBEAPRSP-----APSEGAGSDVFDGLGMAAKGLOSLPTHDPSPLO 1110
Db 514 -----SRTPLLSLSATSNNATNCID-----RNGQGHVPVREDSFVQ 550
Qy 1111 RYSEDPVPLPSET--DGYVAVLTCSPQEPYVNPQDVRPQPPSPREGPLPAARPAGATLE 1168
Db 551 RYSSDPTGNPLESIDDGFL-----PAPEYVNVQ--LMPKKPS----- 585
Qy 1169 RAKTLPSPGKGVVVDVF-----AFGGAIVENPEYLTPOGGAAPQHPPPAP 1213
Db 586 ----TAMVQIQIYNNISLTAISKLPMSRYQNSHSTAVDNPYL-----NTNQSPLA 633
Qy 1214 SPADFNLVYWDQ-----DPPE-----RCAPPSTFKGTPTAENPEYLGLDVP 1254
Db 634 KTFPESPPYIQSGNHQINLDNPDYQDFLPNETKPNGLLKVPAARENPEYLRVAAP 689

RESULT 12
TVYUHV
C;Species: avian erythroblastosis virus
C;Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C;Accession: A00644; A38022
R;Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A;Title: The erbB gene of avian erythroblastosis virus is a member of the src gene fami
A;Reference number: A00644; MUID:84026539; PMID:6313229
A;Accession: A00644
A;Molecule type: DNA
A;Residues: 1-604 <YAW>
A;Cross-references: GB:K01216; NID:G209676; PIDN:AAA42400.1; PID:G209678
R;Debuire, B.; Henry, C.; Benaisa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Martin,
Science 224, 1456-1459, 1984
A;Title: Sequencing the erba gene of avian erythroblastosis virus reveals a new type of
A;Reference number: A38022; MUID:84223957; PMID:6328658
A;Accession: A38022
A;Molecule type: DNA
A;Residues: 1-28 'W', 30-139, 'F', 141-145, 'V', 147-152 <DEB>
A;Cross-references: GB:K02006
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific
F;130-395/Domain: protein kinase homology <KIN>
F;138-146/Region: protein kinase ATP-binding motif
F;165/Active site: Lys #status predicted

Query Match 25.0%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 4.1e-63;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;
Qy 587 CAHYKDPFCVAPCGSVGVKPLDLSYMPIWPKFDEGACQCPINCTHSCVDLDDKGPRAEQ 646
Dy 3 CAHFDGPHCVKACPAVGLGENDTL-VKRYADANAVCOLCHPNCTRGCKGPLEGCP--- 58
Qy 647 RASPLTSTVSAVV-GILLVVVLGVFGILLIKERQOKRKRYTMRLLQETELVEPLTSGA 705
Dy 59 NGSKTPTSAAGVGGGLCLVVGIGLYLRR-HIVKRTLRLLQERLEIVELTSPGE 117
Qy 706 MPNOAQRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKLRENTSPKANK 765
Dy 118 APNOAHLRLKETEFKKVKVLGSGAFGIYKGLWPEGEVKIPVAIKELRENTSPKANK 177
Qy 766 EILDEAYMAGVSGPYVRLGICLTSTVQLVLTMPYVGLLDHVNRNRLGSGODLLNW 825
Dy 178 EILDEAYMASVDNPHVCRLLGICLTSTVQLITLMPYVGLLDVIREHKONIGSQYLLNW 237
Qy 826 CMOIAKMSYLEVRLVHRDLAARNVLKSNHVKITDPGLARLLDDETEYHADGGKVP 885
Dy 238 CVQIAKGNLYLEERLRLVHRDLAARNVLKTPQHVKITDPGLAKLLGADKEKYHAEGGKVP 297
Qy 886 IKMMALESILRRRFTHQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPQ 945
Dy 298 IKMMALESILHRIYTHQSDVMSYGVYVWELMTFGSKPYDGIIPASEISSVLEKGERLPPQ 357
Qy 946 ICTIDVYIMVKWIDSECRPRFRELVSERWARDPQRFVITQ-NEDLGPASPLSTF 1004
Dy 358 ICTIDVYIMVKWIDADSRPKRELTAERFSEKWARDPPRYLVITQDERMELPSPTDSKF 417
Qy 1005 YRSILLEDDMDGLDAEYLYVPOGFFCPDPAPGAGGVNHRHSSSTRSGGDLTLGLE 1064
Dy 418 YRTLWEEEDMEDIVDADYLYVPHOGFF-----NSPST----- 449
Qy 1065 PSEBEAPRSP-----APSEGAGSDVFDGLGMAAKGLQSLPHTDPSLQRYSEDPTVP 1119
Dy 450 -----SRTPLLSLSATSNNATNCID-----RNGQGHVPVREDSFVQRYSSDPTGN 495
Qy 1120 LPSET-DGYVAPITCSPPRYVNPQDVRPQPPSPREGPLPAAPAGATLERAKTLSPGK 1177
Dy 496 FLEESIDGFL-----PAPEYVQ--LMPKKPSTAM----- 524
Qy 1178 NGVVKDPAF-----GAVENPEYLTQGGAAQPPHPPAPSPAFD 1218
Dy 525 --VQNYNFISLTAISKLPMDRSYQNSHSTAVONPEYL-----NTNOSPLAKTVFE 574
Qy 1219 NLYYWDQDPPRGAPPSTFKGTPTAENPEY 1248
Dy 575 SSPYWIQSGNHQ-----INLDNPDY 594

RESULT 13

GOFFE
epidermal growth factor receptor - fruit fly (Drosophila melanogaster)
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C:Species: Drosophila melanogaster
C:Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 11-Jun-1999
C:Accession: A00640; A38021
R:Liivneh, E.; Glazer, L.; Segal, D.; Schlessinger, J.; Shilo, B.Z.
Cell 40, 599-607, 1985
A:Title: The Drosophila EGF receptor gene homolog: conservation of both hormone binding
A:Reference number: A00640; MUID:85124611; PMID:2982499
A:Accession: A00640
A:Molecule type: DNA
A:Residues: 1-1330 <LIV>
A:Cross-references: EMBL:K03054
R:Wadsworth, S.C.; Vincent III, W.S.; Bilodeau-Wentworth, D.
Nature 314, 178-180, 1985
A:Title: A Drosophila genomic sequence with homology to human epidermal growth factor re
A:Reference number: A38021; MUID:85137938; PMID:2983232
A:Accession: A38021

A:Molecule type: DNA
A:Residues: 'A', 832-866, 'V', 868-943, 'QTPSLVK' <WAD>
A:Cross-references: EMBL:X02293; NID:97922; PIN:CAA26157.1; PID:9929565
C:Comment: This sequence is tentative because the introns have not been identified.
C:Genetics:
A:Gene: FlyBase:Egfr
A:Cross-references: FlyBase:FBgn0003731
A:Map position: 2 57F
A:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosph
F:1-732/Domain: extracellular #status predicted <EXT>
F:733-764/Domain: transmembrane #status predicted <TM>
F:765-1330/Domain: intracellular #status predicted <INT>
F:808-1072/Domain: protein kinase homology <KIN>
F:816-924/Region: protein kinase ATP-binding motif
F:1122,300,324,363,518,688,695,700/Binding site: carbohydrate (Asn) (covalent) #status p
F:774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:843/Active site: Lys #status predicted
F:1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predic
Query Match 24.3%; Score 1654.5; DB 1; Length 1330;
Best Local Similarity 30.1%; Pred. No. 8.6e-61;
Matches 415; Conservative 179; Mismatches 414; Indels 371; Gaps 40;
Qy 80 VQGVLIARHQVRQVPLQRLIRIVRGTOLF-----EDNYALAVLDNGDPLNNQYIKANSKF 134
Dy 38 ITNYVIVGLDLIPCTLSYRLQILIRGRTLFSLSVEEEKYALFV-----TYSK----- 83
Qy 135 IGTELQRLSLEILKGGVLIQRNPOLCYQDTILWKDIFHNQOLATLIDTNRSRACHP 194
Dy 84 --MYTLEIPDLRDVLNGVGFHNHNYNLCHMRTIQWSEIVSNGTIDAYNYDFTAPERECPK 141
Qy 195 CSPCKGSRCWGESSEDCOSLTRTVACAGCA--RCKGPLPTCCHEOCACAGCTGPKHSDC 252
Dy 142 CHSCTHG-CWGEFPKNCQKFSKLTSPQACAGRCYGPKECHLFCAGCTGTOKDC 200
Qy 253 LACLHFNHSGICEILHCPALVYNTDTPESMPNPEGRYTFGASCVTACPNYVLSLDVSGCT 312
Dy 201 IACKNFFDEAVSKEECPMKRYNPTTVVLETPGKYAYGATCVKECP-GHLLRDNGACV 259
Qy 313 LVCPHLNQEYTAEDGTORCEKSKPCARVCYGLGMEHLREVRVATSNIOEFACKKIFG 372
Dy 260 RSCPQDKMDKGGG-----CVPCNGPCPKTCPGVTVLH-----AGNIDSFRNCTVIDG 306
Qy 373 SLAFLPSSFDG--DPASNTA-----PQEPQLOVRETLEITGYLYISAWPDSLPLSV 424
Dy 307 NIRILDQTFSGFQDVYANYTNGPRYIPDPERREVFTVKREITGYLNIETHGTFQRLSY 366
Qy 425 FQNLQVIRGILHNGAY-SLTLOGLGISWGLSLRSLRELGSGLALIHHTHLCFVHTVPWD 483
Dy 367 FRNLETHGRQLMESMEFALAIVKSSLYSLEMRLKQISSGVVIOHNRDLCYVSNIRWP 426
Qy 484 QLFNPHQALLHTANRPEDEC-----
Dy 427 AIOKEPEQKVMVNVENLRADLCGKFLTILISVQHNIIMHIFAICREKWNHLLIGSVQRGLL 486
Qy 505 ----- 504
Dy 487 GSWHGSVPYLOELQFQWHLHRLRLWLYIQVINSINQDQKSNEHQLTDACYSVPSVPSLTIER 546
Qy 505 -----VGEGLA-- 510
Dy 547 ARYAIQAGLAMELEQITARSASMRHSKTLPAEGRQVPRWVFLGVCASARAGIAEPLAGR 606
Qy 511 -----CHOLCARGHCWGPGTQCVNCSQFLRGQECVEECRVLQGLPREYV---NARHCLP 562
Dy 607 AVCRKCHPLCELCTNYGYHEQVCSKCTHYKRRQCETEC-----PADHYTDEEQRECFQ 660
Qy 563 CHPECPQNGSVTCTFGPEACQVACAHYK-----DPPF-----CVARCPGSKV-PDL 608
Dy 661 RHPEC-----NG---CTGPGADDCKSCRNFULFDANETGPNYSTNMFNCTSKPLEMRHVNY 714
Qy 609 SYMPIWKFDEEGACQCPINCTHSCVDLDDKGPAPQASPLTSIVSAVVGILLVVVLG 668

Db 715 QYTAIGPY-----CAASPPRSSKITANLD-----NMFIPIITGAVLVPTIC 755
Qy VVFGI-LIKRQOKIRKYT--MRLLOTELVEPLTPSGAMPNOAMRILKETELRKVKV 725
Db 756 ILVVVYICROKQAKKETVKTWALSREDSBPLRFSNIGANICKRIKIVDAELRGV 815
Qy 726 LGSAGFTYVYGIWIPGENYKIPVAIKVLRENTSPKANKILDEAYVMAGVGSPPYSRL 785
Db 816 LGMAGFRVYKGVWPEGENYKIPVAIKELKSTGAESSEBFLREAVIMASEEHVNLKL 875
Qy 786 LGICLTSTVOLVTOIMPYGCILLDHVRENKRLGSDQDILNWCQIAKMSYLEEDVRLVHRD 845
Db 876 LAVCMSSOMMLITQMLPGCLLDVYRNRRDKIGSKALLNWSQIAKMSYLEEERLVRD 935
Qy 846 LAARNVLVK---SPNVKHTDFGLARLDIDETVHADGGKVPKIKWMALESILRRRTHQ 902
Db 936 LAARNVLRLAGEDH----DFGLAKLLSSDSNEYKAAGGKMPKWLALCEIRNVRTSK 991
Qy 903 SDVMSYGVVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYMWIMVKWMD 962
Db 992 SDVWAFGVTIWELLTFQORPHENIPAKDIPDLIEVGLKLEQECISLDIYCTLLSCHWLD 1051
Qy 963 SECPRPRELVSESRWARDQRFVITQNEEDLG--PASPLDSTFYRSLLDD---DMGDL 1017
Db 1052 AAMRPTFKQLTTFVAFARDPGRYLAIGDKFTRLPA-----YTSQDEKDLIRKLAPT 1104
Qy 1018 VDABEYLVPOQGFCCPOPAPCAGGWMVHRHRSSTRSGGDLTLGLRPSBEAP----- 1071
Db 1105 TDGSEATKPDYIQLPKAALGPS-----HRTDCT-----DEMPKLNRYC 1143
Qy 1072 RSLPAPSEGAGSDVFDG---DLGMAAKGLOSPLTHDPSLQRYSEDPTVPLPSETDGYV 1128
Db 1144 KDPNSKNSSTGDDERDSSAREVGVGNLR-----LDLPVDEDDYL 1182
Qy 1129 APLTCSPOPEYVQPDVVRPPSPREGPLPAARFAGATLERAKTLSPGKGVKVDFAFG 1188
Db 1183 MP-TCQPGPNNNNNM-----NPNQNNMAAVGVAAGYM-----DLIGVP 1220
Qy 1189 GAVENPEYL---TPQGAAPQPH-----PPPAFSP-AFDNLXYWD 1224
Db 1221 VSDNPEYLLNAQTLGVGESPIQTGTIGIPVMGGPGTMEVKVPMGPGSEPTSSDHEYND 1279

RESULT 14

S35745

Protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus

C;Species: avian erythroblastosis virus

C;Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997

C;Accession: S35745

R;Vennstrom, B.

submitted to the EMBL Data Library, March 1993

A;Reference number: S35743

A;Accession: S35745

A;Molecule type: DNA

A;Residues: 1-544 <VEN>

A;Cross-references: EMBL:X12707

C;Genetics:

A;Gene: erbB

C;Superfamily: epidermal growth factor receptor; protein kinase homology

C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P

F;135-400/Domain: protein kinase homology <KIN>

F;143-151/Region: protein kinase ATP-binding motif

F;170/Active site: Lys #status predicted

Query Match 24.2%; Score 1647; DB 2; Length 544;

Best Local Similarity 54.9%; Pred. No. 7.3e-61;

Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

Qy 578 GPEADQCVACAHYKDPFCVACPCSGVKPDLISYMPIWKFPDEGACQPCINCHTSCVDL 637

Db 1 GP--DHCMKCAHFIDGPHCVKACPVGLGENDTL-VWKYADANAVCOLCHPNCNTRGCKGP 57

Qy 638 DDGCPAEDASPLTSTIVSAV--GILLVVVLGVVFGILIKRROOKIRKYTMRLLOTEL 696
Db 58 GLEGCP---NGSKTPSTIAAGWGGLLCLVVVGLIGLYLRR--HIVKRKTLRLLQREL 113
Qy 697 VEPITSGAMPNOAMRILKETELRKVKVLGSGAGFTYVYGIWIPGENYKIPVAIKVL 756
Db 114 VEPITSGEAPNAHRIKETELFKVKVGLGFAFTYVYKGLWIPGEKVTIPVAIKELR 173
Qy 757 ENTSPKANKEILDEAYVMAGVGSPPYSRLIGICLTSTVOLVTOIMPYGCILLDHVREN 816
Db 174 EATSPRANKILDEAYVMASVDNPHVCRLLIGICLTSTVOLITQIMPYGCILDIYREHKDN 233
Qy 817 LGSODLLNWCQIAKMSYLEEDVRLVHRDLAARNVLKSNHVKITDFGLARLDIDETE 876
Db 234 IGSQYLLNWCQIAKMSYLEEDVRLVHRDLAARNVLKSNHVKITDFGLARLDIDETE 876
Qy 877 YHAGGGKVPKIKWMALESILRRTHQSDVMSYGVVWELMTFGAKPYDGIPIAREIP 936
Db 294 YHAGGGKVPKIKWMALESILRRTHQSDVMSYGVVWELMTFGSKPYDGIPIAREIP 353
Qy 937 KGERLPQPICTIDVYMWIMVKWMDSECRPRELVSESRWARDQRFVITQNEEDLG 995
Db 354 KGERLPQPICTIDVYMWIMVKWMDSECRPRELVSESRWARDQRFVITQNEEDLG 995
Qy 996 PASPLDSTFYRSLLDDDMGDLVDABEYLVPOQGFCCPOPAPCAGGWMVHRHRSSTRSG 1055
Db 414 LPSPTDSKFTYRLMEBEDMEDIVDABEYLVPHQGF-----NSPST--- 454
Qy 1056 GGDLTGLSESEEARPSPL-----APSEGAGSDVFDGDLGMAAKGLOSPLTHDPSPLQ 1110
Db 455 -----SKTPLLSSLSATSNNSATNCIDRNGG-----H----- 481
Qy 1111 RYSEDPTEVLPSETDGYVAPLTCSPQPEYVQPDVVRPPSPREGPLPAARFAGATLER 1169
Db 482 -----PVREDGFL-----PAPEYVQ--LMPKPESTAMVQIYNYISLTAISK 523
Qy 1170 AKTLSPGKGVKVDVFAFGAVENPEYL 1197
Db 524 LPIDSRYN-----SHSTAVDNPEYL 544

RESULT 15

S00727

kinase-related transforming protein (erbB) (EC 2.7.1.1) - avian erythroblastosis virus

C;Species: avian erythroblastosis virus

C;Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997

C;Accession: S00727

R;Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.

Oncogene Res. 1, 265-278, 1987

A;Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mutant

A;Reference number: S00727; MUID:88217326; PMID:2897102

A;Accession: S00727

A;Molecule type: DNA

A;Residues: 1-545 <SCO>

A;Cross-references: EMBL:X06943

C;Genetics:

A;Gene: erbB

C;Superfamily: epidermal growth factor receptor; protein kinase homology

C;Keywords: ATP; phosphotransferase

F;135-400/Domain: protein kinase homology <KIN>

F;143-151/Region: protein kinase ATP-binding motif

Query Match 24.1%; Score 1640; DB 2; Length 545;

Best Local Similarity 54.9%; Pred. No. 1.4e-60;

Matches 345; Conservative 69; Mismatches 122; Indels 92; Gaps 15;

Qy 578 GPEADQCVACAHYKDPFCVACPCSGVKPDLISYMPIWKFPDEGACQPCINCHTSCVDL 637

Db 1 GP--DHCMKCAHFIDGPHCVKACPVGLGENDTL-VWKYADANAVCOLCHPNCNTRGCKGP 57

Qy 638 DDGCPAEDASPLTSTIVSAV--GILLVVVLGVVFGILIKRROOKIRKYTMRLLOTEL 696

Db 58 GLEGCP---NGSKTPSTIAAGWGGLLCLVVVGLIGLYLRR--HIVKRKTLRLLQREL 113

```
Qy 697 VEPLTPSGAMPNQAOMRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLR 756
Db 114 VEPLTPSGEAPNQAHLRILKETEFKKVKVLGFGAGFTVYKGLWIPEGEKVTIPVAIKELR 173
Qy 757 ENTSPKANKEILDEAYVMAGVSPVSRLLGICLTSTVQLVQVQLMPYGCCLLDHVRENNGR 816
Db 174 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCCLLDYIREHKDN 233
Qy 817 LGSQDLNMCQIAKGMYLEDDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 876
Db 234 IGSQYLLNMCVQIAKGMNYLEERHLVHRDLAARNVLKTPQDVKITDFGLAKQLGADEKE 293
Qy 877 YHADGGKVPICKMALESILRRRFTHOSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLE 936
Db 294 YHAEKKVPICKMALESILHRIYTHOSDVWSYGVTVWELMTFGSKPYDGIPIASEISSVLE 353
Qy 937 KGERLPQPPICTIDVYIMVCKWMIDSECRPRELVSEFSRMDARDPQRFVVIQ-NEDLG 995
Db 354 KGERLPQPPICTIDVYIMVCKWMSDADSRPKFRELIAEFSKMDARDPPRYLVIOGDERMH 413
Qy 996 PASPLDSTFYSLLEDDDMGLVDAAEYLVPOQGFPCPDPAAGAGGWHHRHRSSTRSRG 1055
Db 414 LPSPTDSKFYRLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
Qy 1056 GGDLTLGLEPSEEEAPRSP-----APSEGAGSDVFDGDLGWAAGKGLQSLPTHDPSPQLQ 1110
Db 455 -----SRTPLLSSLSATSNNSATNCIDRNGG-----H----- 481
Qy 1111 RYSEDPTVPLPSETDGYVAPLTCSQPEYVNPQDVVRPQPPSPREGPLPAARPAGAT-LER 1169
Db 482 -----PVREDGFL-----PAPEYVNO--LMPKKPSTAMVQNOIYNYISLTAISK 523
Qy 1170 AKTLSPGKNGVVKDVFEGGAVENPEYL 1197
Db 524 LPMDSRYQN-----SHSTAVDNPEYL 544
```

Search completed: July 22, 2003, 09:08:18
Job time : 30.9062 secs

GenCore version 5.1.6
Copyright (C) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 37.9774 Seconds
(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-210-224-12
Perfect score: 6789
Sequence: 1 MEUAAALRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6704	98.7	1255	21	AAV92620 Human heregulin 2
2	6704	98.7	1255	22	AAE12130 Human tyrosine kin
3	6704	98.7	1255	22	AAE60167 HER2 transgene pla
4	6704	98.7	1255	23	AAU74545 Human HER2 (Erbb2)
5	6698	98.7	1255	17	AAW01111 HER-2/neu protein.
6	6698	98.7	1255	20	AAW92406 Human HER-2/neu on
7	6698	98.7	1255	21	AAE21198 Human HER-2/neu pr
8	6698	98.7	1255	21	AAV84780 Amino acid sequenc
9	6698	98.7	1255	22	AAE85458 Human HER-2/neu pr
10	6698	98.7	1255	22	AAE88267 HER2/neu amino aci

11	6698	98.7	1255	23	AAE24067 Human Her-2 protei
12	6698	98.7	1255	23	AAE20479 Human Her-2/neu pr
13	6698	98.7	1255	23	AAE51143 Human Her-2/neu on
14	6698	98.7	1255	23	AAU77114 Human Her-2/neu po
15	6655	98.0	1433	14	AAE39568 Sequence of c-erbB
16	6534	96.2	1223	23	AAU98923 Human breast cance
17	6381	94.0	1200	21	AAE21208 Human HER-2/neu pr
18	5890.5	86.8	1256	21	AAE21199 Rat HER-2/neu prot
19	5890.5	86.8	1256	23	AAE51144 Rat Her-2/neu onco
20	5865.5	86.4	1256	21	AAE21206 Mouse Her-2/neu pr
21	5865.5	86.4	1256	22	AAE62860 Amino acid sequenc
22	5865.5	86.4	1256	23	AAE51151 Mouse Her-2/neu on
23	4784	70.5	919	21	AAE21203 Human HER-2/neu fu
24	4784	70.5	919	23	AAE51148 Her-2/neu extracel
25	4015.5	59.1	920	23	AAE51152 Mouse Her-2/neu ex
26	4015.5	59.1	926	23	AAE51153 Human HER-2/neu ex
27	3668	54.0	712	21	AAE21204 Human HER-2/neu fu
28	3668	54.0	712	23	AAE51149 Her-2/neu extracel
29	3522	51.9	782	18	AAE19764 Her2-GM-CSF immuno
30	3520	51.8	653	21	AAE21200 Extracellular HER-
31	3520	51.8	653	23	AAE51145 Human ErbB2 oncopr
32	3482	51.3	645	22	AAE60408 Human ErbB2 extrac
33	3482	51.3	645	22	AAE61593 DCBscFV-erbB2EC fu
34	3417	50.3	951	21	AAE44993 Extracellular port
35	3314	48.8	624	11	AAE08222 Human protein for
36	3084	45.4	583	23	AAE20483 Human protein for
37	3084	45.4	587	23	AAE20481 Human protein for
38	3083	45.4	589	23	AAE20484 Human protein for
39	3083	45.4	600	23	AAE20482 Amino acid sequenc
40	3076	45.3	1210	21	AAE19259 Human EGF receptor
41	3076	45.3	1210	21	AAE50616 Human Her-1 protei
42	3076	45.3	1210	23	AAE23019 Human epidermal gr
43	3076	45.3	1210	23	AAE50768 Amino acid sequenc
44	3074	45.3	1210	22	AAE68420 Human epidermal gr
45	3035	44.7	1210	23	AAE51168

ALIGNMENTS

RESULT 1
AA92620
ID AAY92620 standard; Protein; 1255 AA.
XX AAY92620;
XX AC
XX DT 10-AUG-2000 (first entry)
XX DT
XX DE Human heregulin 2 (Her2).
XX KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
XX OS Homo sapiens.
XX FH Key
XX FT Domain
XX FT Location/Qualifiers
FT 1..173 /label= N-terminal
FT 5..25 /note= "mature polypeptide"
FT 5..25 /label= insertion region
FT /note= insertion region
FT 59..73 /label= insertion region
FT /note= insertion region
FT 103..117 /label= insertion region
FT /note= insertion region
FT 149..163 /label= insertion region
FT /note= insertion region
FT 174..323 /label= insertion region

FT Region /label= Cysteine_rich_domain
FT 210..224
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 250..264
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 324..483
FT /label= Ligand_binding_domain
FT 325..339
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 369..383
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 465..479
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 484..623
FT /label= Cysteine_rich_domain
FT 579..593
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 624..654
FT /label= Transmembrane_domain
FT 632..652
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 653..667
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 655..1010
FT /label= Tyrosine_kinase_domain
FT 661..675
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 695..709
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 710..730
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 1011..1235
FT /label= C-terminal_domain
XX WO200020027-A2.
XX
XX 13-APR-2000.
XX
XX 05-OCT-1999; 99WO-DK00525.
XX
XX 05-OCT-1998; 98DK-0001261.
XX 20-OCT-1998; 98US-0105011.
XX
XX (MEBI-) M & E BIOTECH AS.
XX
XX Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
XX Gautam A, Birk P, Karlsson G;
XX WPI; 2000-349917/30.
XX N-PSDB; AAA09455.
XX
XX Inducing immune responses to weakly immunogenic, tumor associated
XX peptide antigens for the treatment of breast and prostate cancer
XX
XX Claim 62; Page 193-198; 220pp; English.
XX
XX This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
XX Her2 can be used in the claimed method as an autovaccine to induce a CTL
XX response. Subdominant CTL epitopes, antibody binding regions and
XX cysteine residues involved in disulfide bonds are preserved in the
XX immunogenized forms. Regions suitable for the insertion of foreign T
XX helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
CC cell-associated peptide antigens (PA) such as those associated with
CC cancers (self-proteins), e.g. human prostate specific membrane antigen
CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
CC The method comprises effecting simultaneous presentation by antigen
CC producing cells (APCs) of the animals immune system of: (1) at least 1
CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
CC B-cell group derived from the cell-associated PA; and (2) at least 1
CC first T-helper cell group which is foreign to the animal. Analogues of
CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
CC part of all known and predicted CTL and B-cell epitopes of the respective
CC PA and including at least one foreign T helper epitope are also claimed.
CC The method is used to treat prostate, prostate/breast or breast cancer
CC when the PA is human PSM, FGF8b and Her2, respectively.
XX
SQ Sequence 1255 AA;

Query Match 98.7%; Score 6704; DB 21; Length 1255;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1242; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASTOVCTGTDMLRLPASPTHLDMLRLHYQGCVQGNL 60
DB 1 MELAALCRWGLLLALLPPGAASTOVCTGTDMLRLPASPTHLDMLRLHYQGCVQGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQGVLIHAHQVROVPLQRLIRVGTOLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIQEVQGVLIHAHQVROVPLQRLIRVGTOLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVTGASPGGLRELQLSLTEILKGGVLIQRNPOLCYQDTILMKDIFHKNNQLA 180
DB 121 DPLNNTTPTVTGASPGGLRELQLSLTEILKGGVLIQRNPOLCYQDTILMKDIFHKNNQLA 180
QY 181 LTLIDNRSRACHPCSPMCKGSRGWESSEDCOSLRTVTCAGGCARCKGKPLQYIKANSK 240
DB 181 LTLIDNRSRACHPCSPMCKGSRGWESSEDCOSLRTVTCAGGCARCKGKPLQYIKANSK 240
QY 241 FIGITELKHSDDLACLHFNHSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
DB 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
QY 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFETLEETGYLIYISAWPDSL 420
DB 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFETLEETGYLIYISAWPDSL 420
QY 421 DLSVFQNLQVIRGRIILHNGAYSILTLQGLISWGLSLRELGSGLALIHNTHLFCFVHTV 480
DB 421 DLSVFQNLQVIRGRIILHNGAYSILTLQGLISWGLSLRELGSGLALIHNTHLFCFVHTV 480
QY 481 PWDQLFRNPQHALLHTANRPEDECVGEGLAACHOLCARGHCWGPGTQCVCNCSQFLRQEC 540
DB 481 PWDQLFRNPQHALLHTANRPEDECVGEGLAACHOLCARGHCWGPGTQCVCNCSQFLRQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPCQPONGSVTCFGEADQCACAHYKDPFCVARC 600
DB 541 VEECRVLQGLPREYVNAHCLPCHPCQPONGSVTCFGEADQCACAHYKDPFCVARC 600
QY 601 PSQVKPDLSPYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCFAEORASPLTISVAVG 660
DB 601 PSQVKPDLSPYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCFAEORASPLTISVAVG 660
QY 661 ILLVVLGVVFGILIKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQMRILKETEL 720
DB 661 ILLVVLGVVFGILIKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQMRILKETEL 720
QY 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKELDEAYVMAGVGP 780
DB 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKELDEAYVMAGVGP 780

Qy 781 YVSRLLGICLTSTVQLVTLQMPYCCLLDHVRENRGRIGSQDLINWCMQIAKMSYLEDDR 840
Db 781 YVSRLLGICLTSTVQLVTLQMPYCCLLDHVRENRGRIGSQDLINWCMQIAKMSYLEDDR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESLRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESLRRFT 900
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKCM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKCM 960
Qy 961 IDSECRPRFRELVSERFARMARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020
Db 961 IDSECRPRFRELVSERFARMARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDDDMGDLVDA 1020
Qy 1021 EYLVPOQGFPCPDPAAGAGMVHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080
Db 1021 EYLVPOQGFPCPDPAAGAGMVHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NQPDVRPQPSPREGPIPAARPAAGATLERAKTLSPGKNGVVKOVFAFGGAVENPEYLTQP 1200
Db 1141 NQPDVRPQPSPREGPIPAARPAAGATLERAKTLSPGKNGVVKOVFAFGGAVENPEYLTQP 1200
Qy 1201 GGAAPQHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPAENPEYLGLDVVPV 1255
Db 1201 GGAAPQHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPAENPEYLGLDVVPV 1255

RESULT 2

AAE12130
ID AAE12130 standard; Protein; 1255 AA.
XX
AC AAE12130;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human tyrosine kinase-type receptor, HER-2.
XX
KW Therapeutic compound; major histocompatibility complex; vaccine;
KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
KW antigen presenting cell; human; tyrosine kinase-type receptor.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Region 774..782
FT /note= "Antigenic epitope"
FT
XX
XX W0200168677-A2.
XX
XX 20-SEP-2001.
XX
XX 16-MAR-2001; 2001WO-US40328.
XX
XX 16-MAR-2000; 2000US-0527487.
XX
XX (GENZ) GENZYME CORP.
XX
XX Nicolette CA;
XX
XX WPI; 2001-616284/71.
DR N-PSDB; AAD19731.
XX
XX Novel synthetic therapeutic compound for inducing immune response and
PT for use in adoptive immunotherapy, has enhanced binding to major
PT histocompatibility molecules and enhanced immunoregulatory properties
PT

XX
PS
XX

Claim 4; Page 63-67; 69pp; English.

The invention relates to synthetic therapeutic compounds (antigenic peptides) with enhanced binding to major histocompatibility complex (MHC) molecules and enhanced immunoregulatory properties relative to their natural counterparts. Compounds of the invention are useful for inducing an immune response in a subject and for use in adoptive immunotherapy. They are useful as components of anti-cancer vaccines and to expand immune effector cells that are specific for cancers characterized by expression of the breast cancer antigen, HER-2. Polynucleotides that encode peptides of the invention are useful as hybridization probes and as primers for the detection of genes of gene transcripts that are expressed in antigen presenting cells (APCs), to confirm transduction of polynucleotides into host cells. The present sequence is human tyrosine kinase-type receptor, HER-2. Compounds of the invention are designed based on the HER-2 antigenic peptide (774-782).

SQ Sequence 1255 AA;

Query Match 98.7%; Score 6704; DB 22; Length 1255;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1242; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLPPGAASQTCTGTDMLRLPASPTHLDMLRHLYQGCVVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASQTCTGTDMLRLPASPTHLDMLRHLYQGCVVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIEVQGVLIHAHQVROVPLQRLIRVGTOLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIEVQGVLIHAHQVROVPLQRLIRVGTOLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQNPOLCYQDITLWKDIFHKNNOLA 180
Db 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQNPOLCYQDITLWKDIFHKNNOLA 180
Qy 181 LTLIDTNRSHACHPCSPMCKGSRGWESSEDCQSLTRTVCAGGCARCKGPLPOYIKANSK 240
Db 181 LTLIDTNRSHACHPCSPMCKGSRGWESSEDCQSLTRTVCAGGCARCKGPLPDCCHQC 240
Qy 241 FTGITELKHSIDCLACLFHNSHIGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSIDCLACLFHNSHIGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPHNOEVTAEQDTCRCKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHNOEVTAEQDTCRCKSKPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEETITGLYIISAWPDSL 420
Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEETITGLYIISAWPDSL 420
Qy 421 DLSVFONLQVIRGRILHNGAYSITLQGLGISWGLRSLRELGSGLALIHNNTHLCFVHTV 480
Db 421 DLSVFONLQVIRGRILHNGAYSITLQGLGISWGLRSLRELGSGLALIHNNTHLCFVHTV 480
Qy 481 PWDOLFERNPHQALLHTANRPEDECYEGGLACHQLCARGHCWGPGTQCNCVSQFLRGQEC 540
Db 481 PWDOLFERNPHQALLHTANRPEDECYEGGLACHQLCARGHCWGPGTQCNCVSQFLRGQEC 540
Qy 541 VEECRVLQGLPREYVNNARHCLPCHPECPQNGSVTCFPGPEADOCVCAHYKDPDFCVAR 600
Db 541 VEECRVLQGLPREYVNNARHCLPCHPECPQNGSVTCFPGPEADOCVCAHYKDPDFCVAR 600
Qy 601 PSGVKPDLSYMPIWKFPDDEEGACQPCPINCTHSCVDLDDKGCFAEQASPLTSIVSAVVG 660
Db 601 PSGVKPDLSYMPIWKFPDDEEGACQPCPINCTHSCVDLDDKGCFAEQASPLTSIVSAVVG 660
Qy 661 ILLVVLGVVFGILLIKRROOKIRKYMRLLOSETVELVEPLTPSGAMPNQAQRILKETEL 720
Db 661 ILLVVLGVVFGILLIKRROOKIRKYMRLLOSETVELVEPLTPSGAMPNQAQRILKETEL 720

Qy 721 RKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSPKANKETILDYAYMAGVSP 780
Db 721 RKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSPKANKETILDYAYMAGVSP 780
Qy 781 YVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRENRLGSGDLLNWCQIAKMSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRENRLGSGDLLNWCQIAKMSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIKWMALESTLRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIKWMALESTLRRFT 900
Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVWKCM 960
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVWKCM 960
Qy 961 IDSECRPRFRELVSERFARMARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDGMDLVDA 1020
Db 961 IDSECRPRFRELVSERFARMARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDGMDLVDA 1020
Qy 1021 EYLVPQGGFCPPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
Db 1021 EYLVPQGGFCPPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAGKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKGVVVKDVFAGGAVENPEYLTPO 1200
Db 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKGVVVKDVFAGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAFSPADNLYYWDQDPPERGAPSTFKGTPTAENPEYGLDVPV 1255
Db 1201 GGAAPQHPHPPAFSPADNLYYWDQDPPERGAPSTFKGTPTAENPEYGLDVPV 1255

RESULT 3
AAB60167
ID AAB60167 standard; Protein; 1255 AA.
XX
AC AAB60167;
XX
DT 03-APR-2001 (first entry)
XX
DE HER2 transgene plasmid construct encoded protein.
XX
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
KW antibody.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200100244-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-US17229.
XX
PR 25-JUN-1999; 99US-0141316.
PR 16-MAR-2000; 2000US-0189844.
XX
PA (GETH) GENENTECH INC.
XX
PI Erickson S, Schwall R;
XX
DR WPI; 2001-061962/07.
DR N-PSDB; AAF24297.
XX
PT Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
PT conjugating the antibody to a maytansinoid -
XX

PS Example 3; Fig 4; 92pp; English.
XX
CC The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ Sequence 1255 AA;

Query Match 98.7%; Score 6704; DB 22; Length 1255;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1242; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLALLPPGAASSTQVCTGDMKRLRSPASETHLDMLRHLRYQGCVVQGNL 60
Db 1 MELAALCRWGLLALLPPGAASSTQVCTGDMKRLRSPASETHLDMLRHLRYQGCVVQGNL 60
Qy 61 ELYTLPTNASLSFLQDIQEVQGVLIHNRQVROVPLQRLIRVGTQLFEDNYALAVLDNG 120
Db 61 ELYTLPTNASLSFLQDIQEVQGVLIHNRQVROVPLQRLIRVGTQLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPLQCYQDTILMKDIFHKNNQLA 180
Db 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPLQCYQDTILMKDIFHKNNQLA 180
Qy 181 LTLIDNTRARACHPCSPMKGRCWGESSEDCOSLTRTYCAGGCARCKGPLQYIKANSK 240
Db 181 LTLIDNTRARACHPCSPMKGRCWGESSEDCOSLTRTYCAGGCARCKGPLQYIKANSK 240
Qy 241 FIGITELKSDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPGRYTFFGASCVTACP 300
Db 241 AGCTGPKISDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPGRYTFFGASCVTACP 300
Qy 301 YNYLSTDVSGSCTLVCPHNLQEVTAEDGTQCEKCKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVSGSCTLVCPHNLQEVTAEDGTQCEKCKPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IQEFACKIKIPGSLAPLPSFDGDPASNTAPLOPEQLQVFEETLEETIGVLIYSAMDPSLP 420
Db 361 IQEFACKIKIPGSLAPLPSFDGDPASNTAPLOPEQLQVFEETLEETIGVLIYSAMDPSLP 420
Qy 421 DLSVFONLQVIRGRILHNGAYSLTLQGLGISWGLSLRELGSGLALIHNNHLCFVHTV 480
Db 421 DLSVFONLQVIRGRILHNGAYSLTLQGLGISWGLSLRELGSGLALIHNNHLCFVHTV 480
Qy 481 PMDQLFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHCWGPGTQCVCNCSQFLRGQEC 540
Db 481 PMDQLFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHCWGPGTQCVCNCSQFLRGQEC 540
Qy 541 VEECRVLQGLPREYVNRHCLPCHPECQPNQSVTCFGEADOCVCAHYKDPFPCVARC 600
Db 541 VEECRVLQGLPREYVNRHCLPCHPECQPNQSVTCFGEADOCVCAHYKDPFPCVARC 600
Qy 601 PSGVKPDLSPYMPIWKPDEEGACQPCINCTHSCVDLDDKGCPEARSPLTSTIVSAVVG 660
Db 601 PSGVKPDLSPYMPIWKPDEEGACQPCINCTHSCVDLDDKGCPEARSPLTSTIVSAVVG 660
Qy 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Qy 721 RKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSPKANKETILDYAYMAGVSP 780
Db 721 RKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSPKANKETILDYAYMAGVSP 780
Qy 781 YVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRENRLGSGDLLNWCQIAKMSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRENRLGSGDLLNWCQIAKMSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIKWMALESTLRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIKWMALESTLRRFT 900

Db -841 LVHRDLAARNVLKSPNHVKTDFGLARLLDIDETEVHADGGKVPKWKMALESILRRRT 900
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVWCWM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVWCWM 960
Qy 961 IDSECRPRFRELVSERFMRDPPQFVVIQNEIDLGPASPLDSTPYRSLLEDMDGLDVA 1020
Db 961 IDSECRPRFRELVSERFMRDPPQFVVIQNEIDLGPASPLDSTPYRSLLEDMDGLDVA 1020
Qy 1021 EYLVPQGGFCFDPAPGAGWVHRHRSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Db 1021 EYLVPQGGFCFDPAPGAGWVHRHRSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Qy 1081 AGSDVPDGLGMGAAGKLSLTHDPSPLQRYSDPTVPLPSETDGYVAPLTCSPPQPEYV 1140
Db 1081 AGSDVPDGLGMGAAGKLSLTHDPSPLQRYSDPTVPLPSETDGYVAPLTCSPPQPEYV 1140
Qy 1141 NOPDVRPQPSREGPIPAARPAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPQ 1200
Db 1141 NOPDVRPQPSREGPIPAARPAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPQ 1200
Qy 1201 GGAAPQHPHPPAFSPAFDNLYYDQDPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255
Db 1201 GGAAPQHPHPPAFSPAFDNLYYDQDPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255

RESULT 4

AAU74545
ID AAU74545 standard; Protein; 1255 AA.

AC AAU74545;

XX 23-APR-2002 (first entry)

XX Human HER2 (ErbB2) polypeptide.

XX Human; HER2; ErbB; epidermal growth factor receptor; receptor;
KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
KW glial disorder; astrocytal disorder; hypothalamic disorder;
KW glandular disorder; macrophagal disorder; epithelial disorder;
KW stromal disorder; blastocoealic disorder; inflammatory disorder;
KW angiogenic disorder; immunological disorder.

XX Homo sapiens.

XX US2002001587-A1.

XX 03-JAN-2002.

XX 16-MAR-2001; 2001US-0811123.

XX 16-MAR-2000; 2000US-189844P.

XX 05-OCT-2000; 2000US-238327P.

XX (ERIC/) ERICKSON S.

XX (SCHW/) SCHWALL R.

XX (SLIW/) SLIKOWSKI M.

XX Erickson S, Schwall R, Sliwowski M;

XX WPI; 2002-163686/21.

XX N-PSDB; ABK14058.

XX Treating tumour characterised by overexpression of epidermal growth
PT factor receptor, ErbB or cancer in mammal, comprises administering
PT anti-ErbB antibody-maytansinoid conjugate to the mammal -

XX Example 3; Fig 7; 93pp; English.

XX The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor
CC receptor (ErbB) and does not respond or responds poorly, to treatment
CC with an anti-ErbB antibody, comprising administering to the mammal an
CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
CC prostate and bladder, preferably breast cancer. The breast cancer is a
CC metastatic breast cancer or an aggressive form of metastatic breast
CC cancer which overexpresses ErbB2. The method is also useful for treating
CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
CC epithelial stromal, blastocoealic, inflammatory, angiogenic and
CC immunological disorders. This sequence represents the human HER2 (ErbB2)
CC polypeptide of the invention.

XX SQ Sequence 1255 AA;

Query Match 98.7%; Score 6704; DB 23; Length 1255;

Best Local Similarity 99.0%; Pred. No. 0;

Matches 1242; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLIALLPPGAASQTCTGTDMLRLPASPEHDLMLRLHLYQGCVVQGNL 60

Db 1 MELAALCRWGLLIALLPPGAASQTCTGTDMLRLPASPEHDLMLRLHLYQGCVVQGNL 60

Qy 61 ELYLPTNASLSFLQDIQEVQGVYLIAHNQVROVPLQRLIRVGTOLFEDNYALAVLDNG 120

Db 61 ELYLPTNASLSFLQDIQEVQGVYLIAHNQVROVPLQRLIRVGTOLFEDNYALAVLDNG 120

Qy 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQNPOLCYODTILAKDIIFHKNNOLA 180

Db 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQNPOLCYODTILAKDIIFHKNNOLA 180

Qy 181 LTLIDTNRSRACHPCSPMCKGSRGWESSEDCOSLTRTVCAGGCARCKGKPLQYIKANSK 240

Db 181 LTLIDTNRSRACHPCSPMCKGSRGWESSEDCOSLTRTVCAGGCARCKGKPLTDCCHCEQC 240

Qy 241 FTIGITELKHSIDCLACLFHNSGICELHCPALVTYNTDTFESMNPREGRTVFGASCVTACP 300

Db 241 AAGCTGPKHSIDCLACLFHNSGICELHCPALVTYNTDTFESMNPREGRTVFGASCVTACP 300

Qy 301 YNYLSTDVSGSCTLVCPHNOEVTAEQTCCKSPCARVCYGLGMEHLREVRVTSAN 360

Db 301 YNYLSTDVSGSCTLVCPHNOEVTAEQTCCKSPCARVCYGLGMEHLREVRVTSAN 360

Qy 361 IQEFACGCKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFTELEETIGVLIYSAPDLSL 420

Db 361 IQEFACGCKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFTELEETIGVLIYSAPDLSL 420

Qy 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLRELGSGLALIHNNTHLCFVHTV 480

Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLRELGSGLALIHNNTHLCFVHTV 480

Qy 481 PWDQLFRNPHQALLHTANRPEDECVEGLACHQLCARGHCWGPGPTQCVNCSOFLRGQEC 540

Db 481 PWDQLFRNPHQALLHTANRPEDECVEGLACHQLCARGHCWGPGPTQCVNCSOFLRGQEC 540

Qy 541 VEECRVLQGLPREYVNRHCLPCHPBCQONGSVTCFGEADOCVCAHYKDPFPCVARC 600

Db 541 VEECRVLQGLPREYVNRHCLPCHPBCQONGSVTCFGEADOCVCAHYKDPFPCVARC 600

Qy 601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCFAEQASPLTISIVSAVVG 660

Db 601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCFAEQASPLTISIVSAVVG 660

Qy 661 ILLVVLGVVFGILIKRROOKIRKYTMRLLOSTELVEPLTPSGAMPNQAQMRILKETEL 720

Db 661 ILLVVLGVVFGILIKRROOKIRKYTMRLLOSTELVEPLTPSGAMPNQAQMRILKETEL 720

Qy 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780

Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780

Qy 781 YVSRLLIGICLTSTVQLVTQIMPYGCILLDHVRENRLGSGQDILLNWCQIAKGSYLEDVR 840

781 YVSRLLGICLTSTVQLVTLMPYGLDHDVRENHGRGLSGDQLLNWCQIAKGSYLEDVR 840
841 LVHRDLAARNVVKSPNHVKITDGLARLADIDETEHADGKVPKIKMALESILRRFT 900
841 LVHRDLAARNVVKSPNHVKITDGLARLADIDETEHADGKVPKIKMALESILRRFT 900
901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVKWM 960
901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVKWM 960
961 IDSECRPRFRELSEFSEFARMARDPQRFVVIQNEDELGPASPLDSTFYRSLLEDGMDLDA 1020
961 IDSECRPRFRELSEFSEFARMARDPQRFVVIQNEDELGPASPLDSTFYRSLLEDGMDLDA 1020
1021 EYLVPQGFPCPDPAAGAGMWHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
1021 EYLVPQGFPCPDPAAGAGMWHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
1081 AGSDVFDGLGMGAAGLQSLPTHDPSPLOKRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
1081 AGSDVFDGLGMGAAGLQSLPTHDPSPLOKRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVAFGGAVENPEYLTQP 1200
1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVAFGGAVENPEYLTQP 1200
1201 GGAAPQHPHPPAPSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255
1201 GGAAPQHPHPPAPSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255

RESULT 5

AAW01111 standard; Protein; 1255 AA.
AAW01111;
01-JAN-1997 (first entry)
HER-2/neu protein.
HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
breast cancer; ovary cancer; colon cancer; lung cancer;
prostate cancer; immunisation; tumour; vaccine; vector.
Homo sapiens.
Key Location/Qualifiers
Domain 676..1255
/label= Intracellular domain
/note= "Claimed domain, useful for immunisation"
WO9630514-A1.
03-OCT-1996.
28-MAR-1996; 96WO-US01689.
31-MAR-1995; 95US-0414417.
(UNIW) UNIV WASHINGTON.
Cheever MA, Disis ML;
WPI; 1996-455361/45.
N-PSDB; AAT40739.
DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
treatment of malignancies with which the HER-2/neu oncogene is
associated
Claim 2; Page 56-61; 71pp; English.

xx Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
cc the product of the HER-2/neu oncogene (see also AAT40739). The
cc protein is over-expressed in various cancers, including breast,
cc ovarian, colon, lung and prostate. The intracellular domain of the
cc protein can be used to immunise an animal against a malignancy with
cc which the oncogene is associated. The polypeptide can be produced
cc in transformed host cells for use in immunisation. Alternatively,
cc animal cells are transfected in vivo or ex vivo with a viral vector
cc that directs expression of the polypeptide.
xx
SQ Sequence 1255 AA;
Query Match 98.7%; Score 6698; DB 17; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLRLPASPTHLDMLRHLYQGQVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLRLPASPTHLDMLRHLYQGQVQGNL 60
QY 61 ELTYLPTNASLSFLODIOEVQGVLIHAHQVQVPLQRLRIVRGTOLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLODIOEVQGVLIHAHQVQVPLQRLRIVRGTOLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGTASPGGLRELQLRLSTEILKGGVLIQRPOLCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTTPTVGTASPGGLRELQLRLSTEILKGGVLIQRPOLCYQDTILWKDIFHKNNQLA 180
QY 181 LTLIDNTRSRACHPCSPMCKGSRGWESSEDQSLTRTVCAAGCARCKGPLPOYIKANSK 240
Db 181 LTLIDNTRSRACHPCSPMCKGSRGWESSEDQSLTRTVCAAGCARCKGPLPDDCCEQC 240
QY 241 FTGITELKHSDDLCLAHFNHSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTAC 300
Db 241 AAGCTGPKHSDDLCLAHFNHSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTAC 300
QY 301 YNYLSTDVGSCTLVCPHNLQEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHNLQEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
QY 361 IQEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOEQVFLEETLEETIGLYISAWPDSL 420
Db 361 IQEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOEQVFLEETLEETIGLYISAWPDSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSILTQGLIGLSWGLSLRELGLSLIHHNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLIGLSWGLSLRELGLSLIHHNTHLCFVHTV 480
QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGTQCVCNCSQFLRGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGTQCVCNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPECQFQNGSVTCFGEADQCAHAKYDPPFCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECQFQNGSVTCFGEADQCAHAKYDPPFCVARC 600
QY 601 PSGVKPDLSPYMPIWKEPDEGACQPCPINCTHSCVDLDDKGCFAEORASPLTISVAVNG 660
Db 601 PSGVKPDLSPYMPIWKEPDEGACQPCPINCTHSCVDLDDKGCFAEORASPLTISVAVNG 660
QY 661 ILLVVLGVVVFGLIKRROOKIRKYTMRLLOQETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVLGVVVFGLIKRROOKIRKYTMRLLOQETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIIPVAIKVLRNTPSPKANKSILDEAYVAGVSP 780
Db 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIIPVAIKVLRNTPSPKANKSILDEAYVAGVSP 780
QY 781 YVSRLLGICLTSTVQLVTLMPYGLDHDVRENHGRGLSGDQLLNWCQIAKGSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTLMPYGLDHDVRENHGRGLSGDQLLNWCQIAKGSYLEDVR 840

Db 1021 EYLVPQGFCDPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPOPEYV 1140
Db 1081 AGSDVFDGDLGMAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPOPEYV 1140
Qy 1141 NQDVRPQPPSPREGPLPAARPAATLERAATLSPGKNGVVKDVFAGGAVENPEYLTPO 1200
Db 1141 NQDVRPQPPSPREGPLPAARPAATLERAATLSPGKNGVVKDVFAGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAFSPADNLYWQDPPERGAPSTPKGTPTAENPYLGLDVPV 1255
Db 1201 GGAAPQHPHPPAFSPADNLYWQDPPERGAPSTPKGTPTAENPYLGLDVPV 1255

RESULT 9
AAB85458
ID AAB85458 standard; Protein; 1255 AA.
AC AAB85458;
DT 25-SEP-2001 (first entry)
DE Human HER-2/neu protein.
DE Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KW oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
OS Homo sapiens.
XX WO200153463-A2.
XX 26-JUL-2001.
XX 19-JAN-2001; 2001WO-US01850.
XX 21-JAN-2000; 2000US-0177545.
XX (CORI-) CORIXA CORP.
XX Cheever MA, Hand-Zimmermann S;
XX WPI; 2001-476112/51.
XX N-PSDB; AAH23392.
XX New antigen-presenting cells, useful as vaccines for eliciting or
XX enhancing an immune response to HER-2/neu protein, particularly useful
XX for treating or preventing cancer, e.g. breast cancer -
XX Claim 2; Page 41-46; 49pp; English.
XX The invention provides an isolated antigen-presenting cell, which
XX expresses at least an immunogenic portion of a polypeptide that produces
XX an immune response to HER-2/neu protein. The antigen-presenting cells are
XX useful as vaccines for eliciting or enhancing an immune response to
XX HER-2/neu protein, particularly in treating or preventing malignancies in
XX which the HER-2/neu oncogene is associated. Specifically, these are
XX useful for treating or preventing cancer, e.g. breast cancer, ovarian,
XX colon, lung or prostate cancers. The present sequence represents
XX the human HER-2/neu protein (also known as p185 or c-erbB2).
XX Sequence 1255 AA;
Query Match. 98.7%; Score 6698; DB 22; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
Qy 1 MELAAICRWGLLALLPPGAASVCTGTDKMLRLPASPTHLDMLRHLHYGCGVQGNL 60
Db 1 MELAAICRWGLLALLPPGAASVCTGTDKMLRLPASPTHLDMLRHLHYGCGVQGNL 60
Qy 61 ELTYLPTNASLFLQDIOEVQGVLIHNRQVPLQRLIRVGTQLPFDNYALAVLDNG 1200

Db 61 ELTYLPTNASLFLQDIOEVQGVLIHNRQVPLQRLIRVGTQLPFDNYALAVLDNG 120
Qy 121 DPLNNTTPTVGTASPGSLRELQRLSLTEILKGGVLIQORNPOLCVQDTILWKDIFHKKNOLA 180
Db 121 DPLNNTTPTVGTASPGSLRELQRLSLTEILKGGVLIQORNPOLCVQDTILWKDIFHKKNOLA 180
Qy 181 LTLIDTNRGRACHPCSPMCKSGRCMGESSEDQSLTRTVTCAGGCARCKGPLPOYIKANSK 240
Db 181 LTLIDTNRGRACHPCSPMCKSGRCMGESSEDQSLTRTVTCAGGCARCKGPLPPTDCHEQC 240
Qy 241 FTGITEKLKSDCLACHLPHNHSIGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP 300
Db 241 AAGCTGPKESDCLACHLPHNHSIGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP 300
Qy 301 YNYLSTDVSGCTLVCPHNOEVTAEQTCCKSKPCARVCYGLMEHLREVRVTSAN 360
Db 301 YNYLSTDVSGCTLVCPHNOEVTAEQTCCKSKPCARVCYGLMEHLREVRVTSAN 360
Qy 361 IOEFAGCKKI FGSIAFLPESFDGDPASNTAPLOEQLVFETLEEITGYLYISAMPDLSL 420
Db 361 IOEFAGCKKI FGSIAFLPESFDGDPASNTAPLOEQLVFETLEEITGYLYISAMPDLSL 420
Qy 421 DLSVFONLQVIRGRILHNGAYSILTQGLGISWGLSLRSLRELGSGLALIIHNTHLCFVHTV 480
Db 421 DLSVFONLQVIRGRILHNGAYSILTQGLGISWGLSLRSLRELGSGLALIIHNTHLCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAGHOLCARGHCWGPGTQCVCNCSOFLRGOEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAGHOLCARGHCWGPGTQCVCNCSOFLRGOEC 540
Qy 541 VEECRVLQGI PREYVNRHCLPCHPECQPNQSVTCFGEADQCVACAHYKDPFCVVARC 600
Db 541 VEECRVLQGI PREYVNRHCLPCHPECQPNQSVTCFGEADQCVACAHYKDPFCVVARC 600
Qy 601 PSGVKPDL SYMPIWKFPDEGACQPCINCTHSCVDLDDKGCAPAEORASPLTSI SA VVG 660
Db 601 PSGVKPDL SYMPIWKFPDEGACQPCINCTHSCVDLDDKGCAPAEORASPLTSI SA VVG 660
Qy 661 ILLVVLGVVFGILIKRQOKIRKYTMRLLOQTELVEPLTPSGAMPNOAMRILKETEL 720
Db 661 ILLVVLGVVFGILIKRQOKIRKYTMRLLOQTELVEPLTPSGAMPNOAMRILKETEL 720
Qy 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKSILDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKSILDEAYVMAGVGP 780
Qy 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGLSGODLLNWCQIAKGSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGLSGODLLNWCQIAKGSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRRFT 900
Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTIDVYIMVKCWM 960
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTIDVYIMVKCWM 960
Qy 961 IDSECRPRFELVSEFSRWARDQRFVVTQNEOLGSPASPLDSTFYRSLLEDGMDLVDA 1020
Db 961 IDSECRPRFELVSEFSRWARDQRFVVTQNEOLGSPASPLDSTFYRSLLEDGMDLVDA 1020
Qy 1021 EYLVPOQGFCDPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
Db 1021 EYLVPOQGFCDPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPOPEYV 1140
Db 1081 AGSDVFDGDLGMAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPOPEYV 1140
Qy 1141 NQDVRPQPPSPREGPLPAARPAATLERAATLSPGKNGVVKDVFAGGAVENPEYLTPO 1200

Db 961 IDSECRPRFRELSEFGRMDPQRFVVIQNEGLPASPLDSTFYRSLLEDMDGLVDA 1020
QY 1021 EEYLVPOQGFPCPDAPGAGGWHHRSSSTRSGGDLTLGLEPSEEEAPRGLAPSEG 1080
Db 1021 EEYLVPOQGFPCPDAPGAGGWHHRSSSTRSGGDLTLGLEPSEEEAPRGLAPSEG 1080
QY 1081 AGSDVFDGLGMAAAGLQSLPHTDPSLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGLGMAAAGLQSLPHTDPSLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NQDVRPQPSREGPLPAAPAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO 1200
Db 1141 NQDVRPQPSREGPLPAAPAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPFPAPFADNLYYWDQDPPERGAPSTFKGTPTAENPEYLGLDVVP 1255
Db 1201 GGAAPQHPFPAPFADNLYYWDQDPPERGAPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 8

AAAY84780

ID AAAY84780 standard; Protein; 1255 AA.

XX

XX

XX

XX 08-AUG-2000 (first entry)

XX

XX Amino acid sequence of the SPLICE erbB-2 receptor protein.

XX

XX SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;
XX tumor cell proliferation; tissue degeneration; arthropathy;
XX bone resorption; inflammatory disease; degenerative disorder;
XX wound healing.

XX

XX Homo sapiens.

XX

XX W0200020579-A1.

XX

XX 13-APR-2000.

XX

XX 01-OCT-1999; 99WO-CA00912.

XX

XX 02-OCT-1998; 98US-0165192.

XX

XX (UYMC-) UNIV MCMASTER.

XX

XX Muller WJ, Siegel PM;

XX

XX WPI; 2000-303768/26.

XX

XX N-PSDB; AAA14812.

XX

XX Nucleic acid encoding an erbB 2 receptor protein designated SPLICE
XX erbB-2, inhibitors of the protein are useful for treatment of cancer -

XX

XX Claim 3; Fig 2; 60pp; English.

XX

XX The present sequence represents a SPLICE erbB-2 receptor protein. The
XX protein has an in-frame deletion of 16 amino acids, 2 of which are
XX conserved cysteine residues, compared to the unspliced protein. The
XX erbB-2 polynucleotide is used to construct probes for detecting
XX disorders of cell transformation such as cancer. Antibodies to the
XX protein may be used to detect SPLICE erbB-2 in a sample. Agents
XX (e.g. antisense oligonucleotides) which inhibit the expression of
XX SPLICE erbB-2 are useful for reducing tumor cell proliferation and
XX treating cancer. Substances which stimulate SPLICE erbB-2 are useful
XX for treating conditions involving damaged cells including conditions
XX in which degeneration of tissue occurs, such as arthropathy, bone
XX resorption, inflammatory diseases, degenerative disorders of the
XX central nervous system and wound healing.

XX

XX Sequence 1255 AA;

XX

SQ

Query Match 98.7%; Score 6698; DB 21; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 1 MELAALCRGILLIALLPPGAASQVCTGCTDMKRLPASPTHLDMLRHLYQGCVVQGNL 60
Db 1 MELAALCRGILLIALLPPGAASQVCTGCTDMKRLPASPTHLDMLRHLYQGCVVQGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQGVLIHAHVNOVQPLQRLIRVRGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQGVLIHAHVNOVQPLQRLIRVRGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKKNQLA 180
Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKKNQLA 180
QY 181 LTLIDNRRSRACHPCSPCKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLPOYIKANSK 240
Db 181 LTLIDNRRSRACHPCSPCKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLPOYIKANSK 240
QY 241 FIGITELKHSDDLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
QY 361 IOEPAGCKKIFGSLAPLPESFGDDPASNTAPLOPEQLQVFETLEETITGLYISANPDSLP 420
Db 361 IOEPAGCKKIFGSLAPLPESFGDDPASNTAPLOPEQLQVFETLEETITGLYISANPDSLP 420
QY 421 DLSVFONLQVIRGRILHNGAYSILTQGLIGISWGLSRSLRELGSGLAIHHNTHLCFVHTV 480
Db 421 DLSVFONLQVIRGRILHNGAYSILTQGLIGISWGLSRSLRELGSGLAIHHNTHLCFVHTV 480
QY 481 PWDQLFRNPQHALLHTANRPEDECVGEGLAHQLCARGHCWGPGTQCVCNCSQFLRGQEC 540
Db 481 PWDQLFRNPQHALLHTANRPEDECVGEGLAHQLCARGHCWGPGTQCVCNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREYVNAHCLFCHPECCQNGSVTCFGPEADQCACAHYKDPFPFCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLFCHPECCQNGSVTCFGPEADQCACAHYKDPFPFCVARC 600
QY 601 PSGVAPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCAPAEORASPLTSIISAVVG 660
Db 601 PSGVAPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCAPAEORASPLTSIISAVVG 660
QY 661 ILLVVVLGVVFGILIKRQCKIRKYTMRLLOETELVEPLTPSGAMPNOAQRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRQCKIRKYTMRLLOETELVEPLTPSGAMPNOAQRILKETEL 720
QY 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVRENTSPKANKEILDEAYVMAGVGSF 780
Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVRENTSPKANKEILDEAYVMAGVGSF 780
QY 781 VYSRLIGLCTSTVOLVTOLMPYGLDHRVNRGRGLSQDILLNMCQIAKMSYLEDVYR 840
Db 781 VYSRLIGLCTSTVOLVTOLMPYGLDHRVNRGRGLSQDILLNMCQIAKMSYLEDVYR 840
QY 841 LVHRLDAARNVLKSPNFKITDFGLARLLDIDETEHADGKVPKIMWALESIILRRRT 900
Db 841 LVHRLDAARNVLKSPNFKITDFGLARLLDIDETEHADGKVPKIMWALESIILRRRT 900
QY 901 HQSDVMSYGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPQPPICITDVTVMYVKWM 960
Db 901 HQSDVMSYGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPQPPICITDVTVMYVKWM 960
QY 961 IDSECRPRFRELVSFSSRMARDPQRFVVIQNEGLPASPLDSTFYRSLLEDMDGLVDA 1020
Db 961 IDSECRPRFRELVSFSSRMARDPQRFVVIQNEGLPASPLDSTFYRSLLEDMDGLVDA 1020
QY 1021 EEYLVPOQGFPCPDAPGAGGWHHRSSSTRSGGDLTLGLEPSEEEAPRGLAPSEG 1080

Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMVWKWM 960
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMVWKWM 960
Qy 961 IDSECRPRFELVSEFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGLDVA 1020
Db 961 IDSECRPRFELVSEFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGLDVA 1020
Qy 1021 EYLVPQGGFFCPDPAAGAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRGLAPSEG 1080
Db 1021 EYLVPQGGFFCPDPAAGAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRGLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAGKQSLPHTDPSFLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGKQSLPHTDPSFLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPQPSREGPLPAARPAGATLERAKTILSPGKGVVVDVFAFGGAVENPEYLTPO 1200
Db 1141 NOPDVRPQPSREGPLPAARPAGATLERAKTILSPGKGVVVDVFAFGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPSTFKGTPTAENPEYLGLDVVP 1255
Db 1201 GGAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 7

AA21198
ID AAB21198 standard; protein; 1255 AA.

XX AC AAB21198;
XX DT 12-JAN-2001 (first entry)

XX DE Human HER-2/neu protein.
XX KW Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
XX KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
XX KW colon cancer.

XX OS Homo sapiens.

XX PN W020004899-A1.

XX PD 03-AUG-2000.

XX PF 28-JAN-2000; 2000WO-US02164.

XX PR 29-JAN-1999; 99US-0117976.

XX PA (CORI-) CORIXA CORP.

XX PA (SMIK) SMITHKLINE BEECHAM.

XX PI Cheever MA, Gheysen D;

XX DR WPI; 2000-505976/45.

XX DR N-PSDB; AAA89736.

XX PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins
XX PT useful for vaccinating against breast, ovarian, colon, lung and
XX PT prostate cancers -
XX PS Claim 52; Fig 7; 128pp; English.
XX CC The present sequence is the human HER-2/neu protein. It is a member of
XX CC the tyrosine kinase family of receptor-like glycoproteins and shows
XX CC homology to the epidermal growth factor receptor (EGFR). It probably
XX CC plays a part in cell growth and/or differentiation. The HER-2/neu
XX CC gene is an oncogene. An HER-2/neu fusion protein comprising a
XX CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
XX CC domain may be used to treat or prevent cancer by eliciting or
XX CC enhancing an immune response to the HER-2/neu protein. It may be used
XX CC to treat malignancies such as breast, ovarian, colon, lung and
XX CC prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.
XX SQ Sequence 1255 AA;
Query Match 98.7%; Score 6698; DB 21; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLALLPPGAASQTCTGTDMLRLPASPTHLDMLRLHLYQGCVVQGNL 60
Db 1 MELAALCRWGLLALLPPGAASQTCTGTDMLRLPASPTHLDMLRLHLYQGCVVQGNL 60
Qy 61 ELYLPTNASLFLQDIQEVQGYVLIHQNQVQVPLQRLIRVRGTOLFEDNYALAVLDNG 120
Db 61 ELYLPTNASLFLQDIQEVQGYVLIHQNQVQVPLQRLIRVRGTOLFEDNYALAVLDNG 120
Qy 121 DPLNNTPTVGTASPGGLRELOLRSLTEILKGGVLIQNPOLCYODTILKNDIFHKNNOLA 180
Db 121 DPLNNTPTVGTASPGGLRELOLRSLTEILKGGVLIQNPOLCYODTILKNDIFHKNNOLA 180
Qy 181 LTLIDTNRGRACHPCSPMCKGRCWGESSEDCOSLTRTVTCAGGCARCKGPLPOYIKANSK 240
Db 181 LTLIDTNRGRACHPCSPMCKGRCWGESSEDCOSLTRTVTCAGGCARCKGPLPDDCHEQC 240
Qy 241 FIGITELKSHDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVAVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVAVTSAN 360
Qy 361 IQBFACKKIFGSLAFLPESFDGDPASNTAPLOEQVFEETLEETITGYLISAWPDSL 420
Db 361 IQBFACKKIFGSLAFLPESFDGDPASNTAPLOEQVFEETLEETITGYLISAWPDSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISMGLSLRELGLALIHNNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISMGLSLRELGLALIHNNTHLCFVHTV 480
Qy 481 PWDOLFNRPHCALHTANRPEDECVEGLACHOLCARGHCWGPGTQCVCNSOFLRQEC 540
Db 481 PWDOLFNRPHCALHTANRPEDECVEGLACHOLCARGHCWGPGTQCVCNSOFLRQEC 540
Qy 541 VEBCRVLQGLFREYVNAHCLPCHPECQONGSVTCFGEADOCVCAHYKDPFVCVARC 600
Db 541 VEBCRVLQGLFREYVNAHCLPCHPECQONGSVTCFGEADOCVCAHYKDPFVCVARC 600
Qy 601 PSGVKPDLSTNPIWKFPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSIVSAVVG 660
Db 601 PSGVKPDLSTNPIWKFPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSIVSAVVG 660
Qy 661 ILLVVLGVVFGILIKRQOKIRKYTWRLLOTELVEPLTPSGAMPNQOMILKETEL 720
Db 661 ILLVVLGVVFGILIKRQOKIRKYTWRLLOTELVEPLTPSGAMPNQOMILKETEL 720
Qy 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVSP 780
Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVSP 780
Qy 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGLSGODLLNWCWQIAKMSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGLSGODLLNWCWQIAKMSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRRFT 900
Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMVWKWM 960
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMVWKWM 960
Qy 961 IDSECRPRFELVSEFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGLDVA 1020

QY 341 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900
QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPPICTIDVYMWKWM 960
DB 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPPICTIDVYMWKWM 960
QY 961 IDSECRPRFRELSEFSEMRDQRFVVIQNEIDLGASPLDSTFYRSLLEDGMDGLVDA 1020
DB 961 IDSECRPRFRELSEFSEMRDQRFVVIQNEIDLGASPLDSTFYRSLLEDGMDGLVDA 1020
QY 1021 EYLVPOQGFCDPAPAGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPRPLAPSEG 1080
DB 1021 EYLVPOQGFCDPAPAGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPRPLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDFTVPLPSETDGVVAELTSPQPEYV 1140
DB 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDFTVPLPSETDGVVAELTSPQPEYV 1140
QY 1141 NQDVRPQPPSPREGPLPAARPAATLERAKTILSPGKNGVVKDVFAPGGAVENPEYLTPO 1200
DB 1141 NQDVRPQPPSPREGPLPAARPAATLERAKTILSPGKNGVVKDVFAPGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERGAAPPSTFKGTPTAENPEYLGLDVVP 1255
DB 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERGAAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 6

AAW92406
ID AAW92406 standard; Protein; 1255 AA.
XX AC AAW92406;
XX AC AAW92406;
DT 21-APR-1999 (first entry)
XX Human HER-2/neu oncogene protein.
XX HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
KW malignancy; treatment; tumour.
XX Homo sapiens.
Key Location/Qualifiers
FT Region 676..1255
FT /note= "region which elicits immune response"
PN US5869445-A.
XX
XX 09-FEB-1999.
XX
XX 01-APR-1996; 96US-0625101.
XX
XX 01-APR-1996; 96US-0625101.
PR 17-MAR-1993; 93US-0033644.
PR 12-AUG-1993; 93US-0106112.
PR 31-MAR-1995; 95US-0414417.
XX
XX (UNIW) UNIV WASHINGTON.
XX
XX Cheever MA, Disis ML;
XX WPI; 1999-152835/13.
DR N-PSDB; AAX01912.
XX
XX Use of HER-2/neu polypeptides - for eliciting an immune response to
PT an HER-2/neu associated malignancy, particularly for treating or
PT preventing tumours
XX
XX Claim 3; Column 31-38; 26pp; English.
PS
XX This sequence represents the human HER-2/neu oncogene protein. A fragment

CC of this protein is used in a method for eliciting or enhancing an immune
CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
CC B cells to produce an immune response to the HER-2/neu protein. The
CC method can be used for immunisation against a malignancy in which the
CC HER-2/neu oncogene is associated and in the treatment of an existing
CC tumour, or to prevent tumour occurrence or reoccurrence.
XX

Sequence 1255 AA;

Query Match 98.7%; Score 6698; DB 20; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLRLPASPETHLDMLRLHYLQGVVQGNL 60
DB 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLRLPASPETHLDMLRLHYLQGVVQGNL 60
QY 61 ELYLPTNASLSFLQDIQEVQGVLIHNOVQVPLQRLIRVRGTQLFEDNYALAVLDNG 120
DB 61 ELYLPTNASLSFLQDIQEVQGVLIHNOVQVPLQRLIRVRGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRPOLCYODTILWKDIFHKNNOLA 180
DB 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRPOLCYODTILWKDIFHKNNOLA 180
QY 181 LTLIDTNRGRACHPCSPCKGSRGWESSEDQSLTRTVCAAGCARCKGKPLPOYIKANSK 240
DB 181 LTLIDTNRGRACHPCSPCKGSRGWESSEDQSLTRTVCAAGCARCKGKPLPOYIKANSK 240
QY 241 FTITELKHSDDLACLFHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVCSCTLVCPHNOEVTAEQGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
DB 301 YNYLSTDVCSCTLVCPHNOEVTAEQGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
QY 361 IOEFAGCKKIFGSLAPLESFPGDPAANTAPLOPELVFETLEETITGLYLSAWPDSLP 420
DB 361 IOEFAGCKKIFGSLAPLESFPGDPAANTAPLOPELVFETLEETITGLYLSAWPDSLP 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRELGLSLALIHNNHLCFVHTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRELGLSLALIHNNHLCFVHTV 480
QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAGLACHQLCARGHCWGPPTQVCNCSQFLRGQEC 540
DB 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAGLACHQLCARGHCWGPPTQVCNCSQFLRGQEC 540
QY 541 VEECRVLOGLPREYVNAHCLPCHPECOPOGNSVTCFGEADQCVACAHYKDPFPCVARC 600
DB 541 VEECRVLOGLPREYVNAHCLPCHPECOPOGNSVTCFGEADQCVACAHYKDPFPCVARC 600
QY 601 PSGVKPDLSYMPIWKFPDEBEGACQPCINCHTHSCVDLDDKGPAPQASPLTSIVSAVVG 660
DB 601 PSGVKPDLSYMPIWKFPDEBEGACQPCINCHTHSCVDLDDKGPAPQASPLTSIVSAVVG 660
QY 661 ILLVVVLGVVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNQAQRILKETEL 720
DB 661 ILLVVVLGVVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNQAQRILKETEL 720
QY 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVALKVLRENTSPKANKEILDEAYMAGVGS 780
DB 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVALKVLRENTSPKANKEILDEAYMAGVGS 780
QY 781 YVSRLLGICLTSTVQLVTQLMFYGLLDHVRNRLGSLQDILLNWCMIKAGMSYLEDYR 840
DB 781 YVSRLLGICLTSTVQLVTQLMFYGLLDHVRNRLGSLQDILLNWCMIKAGMSYLEDYR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:05:49 ; Search time 0.755229 Seconds
(without alignments)
1909.378 Million cell updates/sec

Title: US-09-806-703A-12

Perfect score: 74
Sequence: 1 QYIKANSKFIGITEL 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	74	100.0	1315	1 BTCLTN	tentoxilysin (EC 3
2	44.5	60.1	244	2 S29982	class II histocomp
3	44	59.5	66	2 S31029	gene 84 protein -
4	43	58.1	180	2 G86826	diamine N-acetyltr
5	43	58.1	899	2 T42976	hypothetical prote
6	42.5	57.4	1060	2 S06286	major merozoite su
7	42.5	57.4	1086	2 S16752	major merozoite su
8	42.5	57.4	1701	2 A54498	major merozoite su
9	42.5	57.4	1701	2 A26868	major merozoite su
10	42.5	57.4	1726	1 SAZQGM	major merozoite su
11	42.5	57.4	1726	2 A45948	major merozoite su
12	42	56.8	1333	2 S38635	blastoplia polyprot
13	41	55.4	123	2 G48677	Ig heavy chain V-D
14	41	55.4	447	2 H97146	siderophore/Surfac
15	41	55.4	899	2 G36812	hypothetical prote
16	40.5	54.7	245	2 S29980	class II histocomp
17	40	54.1	79	2 D85794	hypothetical prote
18	40	54.1	194	2 G64026	lacyl-carrier-prot
19	40	54.1	601	1 A55485	oligopeptidase (EC
20	40	54.1	601	2 G86840	oligopeptidase
21	40	54.1	644	2 S46746	hypothetical prote
22	39	52.7	102	2 PH1491	Ig heavy chain V r
23	39	52.7	119	2 PH1516	Ig heavy chain V r
24	39	52.7	119	2 PH1518	Ig heavy chain V r
25	39	52.7	119	2 PH1519	Ig heavy chain V r
26	39	52.7	123	2 F48677	Ig heavy chain V-D
27	39	52.7	135	2 PH1494	Ig heavy chain V r
28	39	52.7	140	2 PH1488	Ig heavy chain V r
29	39	52.7	189	2 G97978	conserved hypoten

adenylate kinase (30	39	52.7	213	1 KIYWC
type II restrictio	31	39	52.7	326	2 B71808
probable acetyl-Co	32	39	52.7	349	2 T43043
type IIS restricti	33	39	52.7	423	2 F64690
hypothetical prote	34	39	52.7	505	2 C90569
hypothetical prote	35	38	51.4	188	2 H71875
hypothetical prote	36	38	51.4	188	2 A64639
hypothetical prote	37	38	51.4	256	2 F64472
tRNA-pseudouridine	38	38	51.4	287	2 F70361
probable hexosyltr	39	38	51.4	381	2 F71196
hypothetical prote	40	38	51.4	383	2 T51466
hypothetical prote	41	38	51.4	424	2 T29127
endo-1,4-beta gluc	42	38	51.4	501	2 A86158
cellulase (EC 3.2.	43	38	51.4	501	2 T52135
methyl-accepting c	44	38	51.4	561	2 E82395
hemagglutinin-neur	45	38	51.4	572	1 HNNZ80

ALIGNMENTS

RESULT 1

BTCLTN

tentoxilysin (EC 3.4.24.68) precursor - Clostridium tetani

N;Alternate names: tetanus neurotoxin

C;Species: Clostridium tetani

C;Date: 31-Mar-1988 #sequence revision 31-Mar-1988 #text change 03-Jun-2002

C;Accession: A25689;A25757; A25194; B25194; A60759; S69348; S09364

R;Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, M.,

EMBO J. 5, 2495-2502, 1986

A;Title: Tetanus toxin: primary structure, expression in E. coli, and homology with botu

A;Reference number: A25689; MUID:87053814; PMID:3536478

A;Accession: A25689

A;Molecule type: DNA

A;Residues: 1-1315 <EIS>

A;Cross-references: GB:X04436; NID:940773; PIDN:CAA28033.1; PID:940770

R;Fairweather, N.F.; Lyness, V.A.

Nucleic Acids Res. 14, 7809-7812, 1986

A;Title: The complete nucleotide sequence of tetanus toxin.

A;Reference number: A25757; MUID:87040747; PMID:3774547

A;Accession: A25757

A;Molecule type: DNA

A;Residues: 1-1315 <FAI>

A;Cross-references: GB:X06214; NID:940773; PIDN:CAA29564.1; PID:940774

R;Experimental source: strain CN3911

R;Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.

J. Bacteriol. 165, 21-27, 1986

A;Title: Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in E

A;Reference number: A25194; MUID:86085672; PMID:3510187

A;Accession: A25194

A;Molecule type: DNA

A;Residues: 743-1315 <FA2>

A;Cross-references: GB:M12739; NID:944920; PIDN:AAA23282.1; PID:9144921

A;Accession: B25194

A;Molecule type: protein

A;Residues: 865-894 <FA3>

R;Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.

Infect. Immun. 57, 3588-3593, 1989

A;Title: Isolation, purification, and characterization of fragment B, the NH-2-terminal t

A;Reference number: A60759; MUID:90035436; PMID:2478476

A;Accession: A60759

A;Molecule type: protein

A;Residues: 461-475 <MAT>

R;Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.

J. Immunol. 142, 394-402, 1989

A;Title: Delineation of several DR-restricted tetanus toxin T cell epitopes.

A;Reference number: JS0098; MUID:89093918; PMID:2463305

A;Contents: annotation; epitope region

R;Schlavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R.

Nature 359, 832-835, 1992

A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic

A;Reference number: S27125; MUID:93063293; PMID:1331807

A;Contents: annotation

R;de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.; Montecucco, C.
Eur. J. Biochem. 229, 61-69, 1995
A;Title: Structural studies on the zinc-endopeptidase light chain of tetanus neurotoxin.
A;Reference number: S69348; MUID:95262688; PMID:7744050
A;Accession: S69348
A;Molecule type: protein
A;Residues: 2-31 <DEF>
C;Comment: The source of this protein was an extrachromosomal plasmid.
C;Comment: The precursor is cleaved by endogenous proteinase activity to form light (fragment A) and heavy (fragment B) chains. The amino end of the heavy chain (fragment B) forms ion channels in a lipid bilayer. Fragment C binds to ganglionic presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of synaptobrevin.
A;Description: blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide bond in synaptobrevin.
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-457/Product: tetroxylisin light chain (fragment A) #status predicted <RTL>
F;461-1315/Product: tetroxylisin heavy chain (fragment B.C) #status experimental <TTH>
F;461-864/Domain: channel forming (fragment B) #status predicted <TXB>
F;865-1315/Domain: ganglioside binding (fragment C) #status predicted <TXC>
F;233-237/Binding site: zinc (His) #status predicted
F;234/Active site: Glu #status predicted

Query Match 100.0%; Score 74; DB 1; Length 1315;
Best Local Similarity 100.0%; Pred. No. 3.6e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QYIKANSKFIGITEL 15
Db 830 QYIKANSKFIGITEL 844

RESULT 2
S29982
class II histocompatibility antigen - Atlantic salmon
C;Species: Salmo salar (Atlantic salmon)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Sep-1999
C;Accession: S29982
R;Hardvik, I.
Submitted to the EMBL Data Library, October 1992
A;Reference number: S29980
A;Accession: S29982
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-244 <HOR>
A;Cross-references: EMBL:X70166; NID:964369; PID:964370
C;Superfamily: class II histocompatibility antigen; immunoglobulin homology

Query Match 60.1%; Score 44.5; DB 2; Length 244;
Best Local Similarity 55.6%; Pred. No. 1.8;
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

Qy 1 QYIKANS---KFIGITEL 15
Db 51 EYIRFNSVGVKFGVGTTEL 68

RESULT 3
S31029
gene 84 protein - Mycobacterium phase L5
C;Species: Mycobacterium phase L5
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 08-Oct-1999
C;Accession: S31029
R;Donnelly-Wu, M.K.; Jacobs Jr., W.R.; Hatfull, G.F.
Mol. Microbiol. 7, 407-417, 1993
A;Title: Superinfection immunity of mycobacteriophage L5: applications for genetic transduction.
A;Reference number: S30949; MUID:93211283; PMID:8459767
A;Accession: S31029
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-66 <DON>
A;Cross-references: EMBL:Z18946; NID:915859; PIDN:CAA79460.1; PID:e59702; PID:g579152

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1992
C;Genetics:
A;Gene: 84
A;Start codon: GTG

Query Match 59.5%; Score 44; DB 2; Length 66;
Best Local Similarity 57.1%; Pred. No. 0.58;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 YIKANSKFIGITEL 15
Db 50 YIKRNGKFGVGTWEV 63

RESULT 4
G86826
diamine N-acetyltransferase (EC 2.3.1.57) [imported] - Lactococcus lactis subsp. lactis
N;Alternate names: spermidine acetyltransferase
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 17-May-2002
C;Accession: G86826
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarre, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. lactis
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: G86826
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-180 <STO>
A;Cross-references: GB:AE005176; PID:gl2724622; PIDN:AAK05713.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: ynfF
C;Superfamily: Escherichia coli ribosomal-protein-alanine N-acetyltransferase rimJ
C;Keywords: acyltransferase; coenzyme A

Query Match 58.1%; Score 43; DB 2; Length 180;
Best Local Similarity 69.2%; Pred. No. 2.5;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 3 IKANSKFIGITEL 15
Db 65 IEANDTFIGITEL 77

RESULT 5
T42976
hypothetical protein 63 - ateline herpesvirus 3 (strain 73)
C;Species: ateline herpesvirus 3
A;Variety: strain 73
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C;Accession: T42976
R;Albrecht, J.C.; Fleckenstein, B.
Submitted to the EMBL Data Library, August 1998
A;Description: Primary structure of the herpesvirus ateles genome.
A;Reference number: Z22274
A;Accession: T42976
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-899 <ALB>
A;Cross-references: EMBL:AF083424; PIDN:AAC95587.1
A;Experimental source: strain 73

Query Match 58.1%; Score 43; DB 2; Length 899;
Best Local Similarity 64.3%; Pred. No. 13;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QYIKANSKFIGITE 14
Db 124 QYITSNTFTGQTE 137

RESULT 6

C;Keywords: surface antigen

Query Match 57.4%; Score 42.5; DB 2; Length 1701;
Best Local Similarity 60.0%; Pred. No. 30;
Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

QY 1 QYIKANSKFI-GITE 14
|:::|||||::|
Db 1001 QFVKNSKVITGLTE 1015

RESULT 9

A26868
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (str;
C;Species: Plasmodium falciparum
C;Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 09-Jun-2000
C;Accession: A26868
R;Tanabe, K.; Mackay, M.; Goman, M.; Scaife, J.G.
J. Mol. Biol. 195, 273-287, 1987
A;Title: Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium
A;Reference number: A26868; MUID:88011243; PMID:3079521
A;Accession: A26868
A;Molecule type: DNA
A;Residues: 1-1701 <TAN>
C;Superfamily: major merozoite surface antigen
C;Keywords: surface antigen
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1701/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 57.4%; Score 42.5; DB 2; Length 1701;
Best Local Similarity 60.0%; Pred. No. 30;
Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

QY 1 QYIKANSKFI-GITE 14
|:::|||||::|
Db 1001 QFVKNSKVITGLTE 1015

RESULT 10

SAZOGM
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (str;
N;Alternate names: 195K glycoprotein
C;Species: Plasmodium falciparum
C;Date: 30-Sep-1987 #sequence_revision 31-Mar-1991 #text_change 09-Jun-2000
C;Accession: A23386; S06361
R;Weber, J.L.; Leininger, W.M.; Lyon, J.A.
Nucleic Acids Res. 14, 3311-3323, 1986
A;Title: Variation in the gene encoding a major merozoite surface antigen of the human ma
A;Reference number: A23386; MUID:86205236; PMID:3517809
A;Accession: A23386
A;Molecule type: DNA
A;Residues: 1-1104 <WEB1>
A;Cross-references: EMBL:X03831
R;Weber, J.L.; Sim, B.K.L.; Lyon, J.A.; Wolff, R.
Nucleic Acids Res. 16, 1206, 1988
A;Title: Merozoite surface protein sequence from the Camp strain of the human malaria pa
A;Reference number: S06361; MUID:89143999; PMID:3278296
A;Accession: S06361
A;Molecule type: DNA
A;Residues: 1104-1726 <WEB2>
A;Cross-references: EMBL:X03831
C;Comment: The merozoite stages of different strains have strain-specific surface antigen
C;Comment: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The mer
C;Superfamily: major merozoite surface antigen
C;Keywords: glycoprotein; merozoite; surface antigen; tandem repeat
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1726/Product: major merozoite surface antigen #status predicted <MAT>
F;67-87, 91-96, 100-105, 109-120/Region: 3-residue repeats (S-G-T)
F;757-765/Region: 3-residue repeats (T-E-E)
F;133, 272, 501, 567, 638, 827, 839, 924, 944, 990, 1016, 1114, 1221, 1613, 1658/Binding site: carbohy

Query Match 57.4%; Score 42.5; DB 1; Length 1726;
Best Local Similarity 60.0%; Pred. No. 31;

Oy 1 QYIKANSKFIGITE 14
Db 124 QYITSNATFTGLSE 137

Search completed: July 22, 2003, 08:12:43
Job time : 2.75523 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:05:49 ; Search time 1.51046 Seconds
(without alignments)
2046.206 Million cell updates/sec

Title: US-09-806-703A-12

Perfect score: 74

Sequence: 1 QYIKANSKFIGITEL 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	74	100.0	1310	2 Q93N27	Q93n27 clostridium
2	44.5	60.1	60	7 Q31585	Q31585 salmo salar
3	44.5	60.1	71	7 Q9XRJ9	Q9xrj9 salvelinus
4	44.5	60.1	85	7 Q951S2	Q951s2 salmo salar
5	44.5	60.1	85	7 Q95HY1	Q95hy1 salmo salar
6	44.5	60.1	86	7 Q95HX4	Q95hx4 salmo salar
7	44.5	60.1	244	7 Q31590	Q31590 salmo salar
8	44	59.5	545	10 Q9XG37	Q9xg37 guillardia
9	43.5	58.8	67	7 Q31578	Q31578 salmo salar
10	43	58.1	180	16 Q9CF66	Q9cf66 lactococcus
11	43	58.1	250	9 Q9MCL7	Q9mcl7 streptococ
12	43	58.1	252	9 Q9XJ8	Q9xj8 lactococcus
13	43	58.1	302	11 Q9CRV4	Q9crv4 mus musculu
14	43	58.1	309	11 Q9CYD2	Q9cyd2 mus musculu
15	43	58.1	899	12 Q9YTK4	Q9ytk4 ateline her
16	42.5	57.4	1087	5 Q25961	Q25961 plasmodium

17	42.5	57.4	1694	5 Q9NHX1	Q9nhx1 plasmodium
18	42.5	57.4	1694	5 Q9TZT5	Q9tzt5 plasmodium
19	42.5	57.4	1704	5 Q9TZT4	Q9tzt4 plasmodium
20	42.5	57.4	1720	5 Q25922	Q25922 plasmodium
21	42	56.8	358	17 Q8TZU8	Q8tzu8 pyrococcus
22	42	56.8	1333	5 Q24262	Q24262 drosophila
23	41.5	56.1	84	13 Q9DEK4	Q9dek4 coregonus s
24	41.5	56.1	85	7 Q95IS3	Q95is3 salmo salar
25	41.5	56.1	85	7 Q95IR2	Q95ir2 salmo salar
26	41.5	56.1	149	7 Q31495	Q31495 oncorhynch
27	41.5	56.1	216	7 Q9GJH0	Q9gjh0 salmo trutt
28	41.5	56.1	216	7 Q9GJG9	Q9gig9 salmo trutt
29	41	55.4	187	16 Q8R9Q7	Q8r9q7 thermoanaer
30	41	55.4	247	11 Q9D3B9	Q9d3b9 mus musculu
31	41	55.4	384	4 Q9HD07	Q9hd07 homo sapien
32	41	55.4	384	16 Q8RCF9	Q8rcf9 thermoanaer
33	41	55.4	447	16 Q97HK7	Q97hk7 clostridium
34	41	55.4	532	5 Q96671	Q96671 drosophila
35	41	55.4	540	5 Q9VU53	Q9vu53 drosophila
36	40.5	54.7	67	7 Q31577	Q31577 salmo salar
37	40.5	54.7	67	7 Q31581	Q31581 salmo salar
38	40.5	54.7	67	7 Q31582	Q31582 salmo salar
39	40.5	54.7	71	7 Q9XRH1	Q9xrh1 salvelinus
40	40.5	54.7	71	7 Q9XRH6	Q9xrh6 salvelinus
41	40.5	54.7	84	13 Q9DEJ6	Q9dej6 coregonus s
42	40.5	54.7	85	7 Q95IS9	Q95is9 salmo salar
43	40.5	54.7	85	7 Q95IS8	Q95is8 salmo salar
44	40.5	54.7	85	7 Q95IS6	Q95is6 salmo salar
45	40.5	54.7	85	7 Q95IS5	Q95is5 salmo salar

ALIGNMENTS

RESULT 1

Q93N27 PRELIMINARY; PRT; 1310 AA.

AC Q93N27; 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)

DE Tetanus toxin (Fragment).

OS Clostridium tetani.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;

OC Clostridiales; Clostridiaceae; Clostridium.

OX NCBI_TaxID=1513;

RN [1]

RP SEQUENCE FROM N.A.

RA Shumin Z., Dianliang L.;

RT "Cloning and sequence analysis of tetanus toxin gene.";

RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF389424; AAK72964.2; -.

DR InterPro; IPR000395; Bontoxilysin.

DR InterPro; IPR001064; Crystallin.

DR InterPro; IPR000130; Zn_Mp2ptdee.

DR Pfam; PF01742; Peptidase M27; 1.

DR ProDom; PD001963; Bontoxilysin; 1.

DR PROSITE; PS00225; CRYSTALLIN BETA GAMMA; UNKNOWN_1.

DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.

FT NON_TER 1310 1310

SQ SEQUENCE 1310 AA; 150316 MW; 9EADDC914418E450 CRC64;

Query Match 100.0%; Score 74; DB 2; Length 1310;

Best Local Similarity 100.0%; Pred. No. 0.00019;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15

Db 831 QYIKANSKFIGITEL 845

RESULT 2

```

Q31585
ID Q31585 PRELIMINARY; PRT; 60 AA.
AC Q31585;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE (DB02) MHC class II beta 1 (Fragment).
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RA Grimholt U., Oleaker I., de Vries Lindstrom C., Lie O.;
RT "A study of polymorphism in the MHC class II beta 1 and MHC class I
RT alpha 2 domain exons of Atlantic salmon (Salmo salar).";
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; L24953; AAA49597.1; -.
DR InterPro; IPR000353; MHC_II_beta.
DR Pfam; PF00969; MHC_II_beta.1.
DR ProDom; PD000328; MHC_II_beta.1.
KW Glycoprotein; MHC II; Transmembrane.
FT NON_TER 1
FT NON_TER 60
SQ SEQUENCE 60 AA; 6776 MW; 5D4F3449060940E2 CRC64;

Query Match 60.1%; Score 44.5; DB 7; Length 60;
Best Local Similarity 55.6%; Pred. No. 1.6;
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 1 QYIKANS---KFIGITEL 15
Db 16 EYIRFNSVTGKFGVGYTEL 33

RESULT 3
Q9XRJ9 PRELIMINARY; PRT; 71 AA.
AC Q9XRJ9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MHC class II beta 1 (Fragment).
GN SANA.
OS Salvelinus namaycush (Lake trout).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salvelinus.
OX NCBI_TaxID=8040;
RN [1]
RP SEQUENCE FROM N.A.
RA Dorschner M.O., Duris T., Phillips R.B.;
RT "Diversity of a Lake Trout Mhc class II Gene.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF130026; AAD20889.1; -.
DR InterPro; IPR000353; MHC_II_beta.
DR Pfam; PF00969; MHC_II_beta.1.
DR ProDom; PD000328; MHC_II_beta.1.
KW Glycoprotein; MHC II; Transmembrane.
FT NON_TER 1
FT NON_TER 71
SQ SEQUENCE 71 AA; 8002 MW; E3095286582A9F2D CRC64;

Query Match 60.1%; Score 44.5; DB 7; Length 71;
Best Local Similarity 55.6%; Pred. No. 1.9;
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 1 QYIKANS---KFIGITEL 15
Db 14 EYIRFNSVTGKFGVGYTEL 31

RESULT 4
Q9SIS2 PRELIMINARY; PRT; 85 AA.
ID Q9SIS2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE MHC class II beta chain (Fragment).
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RA Landry C., Bernatchez L.;
RT "Comparative analysis of population structure across environments and
RT geographic scales at Major Histocompatibility Complex and
RT microsatellite in Atlantic salmon (Salmo salar).";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF373699; AAK61882.1; -.
DR InterPro; IPR000353; MHC_II_beta.
DR Pfam; PF00969; MHC_II_beta.1.
DR ProDom; PD000328; MHC_II_beta.1.
KW Glycoprotein; MHC; MHC II; Transmembrane.
FT NON_TER 1
FT NON_TER 85
SQ SEQUENCE 85 AA; 9743 MW; 3214E01AD1B66AC5 CRC64;

Query Match 60.1%; Score 44.5; DB 7; Length 85;
Best Local Similarity 55.6%; Pred. No. 2.2;
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 1 QYIKANS---KFIGITEL 15
Db 33 EYIRFNSVTGKFGVGYTEL 50

RESULT 5
Q9SHY1 PRELIMINARY; PRT; 85 AA.
ID Q9SHY1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE MHC class II B antigen (Fragment).
GN DBI.
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RA Langefors A., Lohm J., von Schantz T.;
RT "Allelic polymorphism in MHC class II B in four populations of
RT Atlantic salmon (Salmo salar).";
RL Immunogenetics 53:329-336(2001).
DR EMBL; AF104370; AAL04002.1; -.
DR InterPro; IPR000353; MHC_II_beta.
DR Pfam; PF00969; MHC_II_beta.1.
DR ProDom; PD000328; MHC_II_beta.1.
KW Glycoprotein; MHC; MHC II; Transmembrane.
FT NON_TER 1
FT NON_TER 85
SQ SEQUENCE 85 AA; 9723 MW; 27C9F7931F1F01C4 CRC64;

Query Match 60.1%; Score 44.5; DB 7; Length 85;
Best Local Similarity 55.6%; Pred. No. 2.2;
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 1 QYIKANS---KFIGITEL 15

```



```
Db          :||: || ||: ||||
33 EYIRFNSTVGKFGVGYTEL 50

RESULT 6
Q95HX4      PRELIMINARY;      PRT;      86 AA.
AC Q95HX4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE MHC class II B antigen (Fragment).
GN DBI.
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21383619; PubMed=11491536;
RA Langefors A., Lohm J., von Schantz T.;
RT "Allelic polymorphism in MHC class II B in four populations of
AT Atlantic salmon (Salmo salar)".
RL Immunogenetics 53:329-336(2001).
DR EMBL; AF104377; AAL04009.1; -.
DR InterPro; IPR000353; MHC_II_beta.
DR Pfam; PF00969; MHC_II_beta; 1.
DR ProDom; PD000328; MHC_II_beta; 1.
KW Glycoprotein; MHC; MHC II; Transmembrane.
FT NON_TER 1 86
FT TER 86
SQ SEQUENCE 86 AA; 9912 MW; E5097729F681F149 CRC64;

Query Match 60.1%; Score 44.5; DB 7; Length 86;
Best Local Similarity 55.6%; Pred. No. 2.2;
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 1 QYIKANS---KFIGITEL 15
: ||: || ||: ||||
33 EYIRFNSTVGKFGVGYTEL 50

Db          :||: || ||: ||||
33 EYIRFNSTVGKFGVGYTEL 50

RESULT 7
Q31590      PRELIMINARY;      PRT;      244 AA.
AC Q31590;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MHC class II.
GN MHC-SASA CLASS II B.
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=LEUKOCYTE;
RX MEDLINE=93170890; PubMed=8436418;
RA Hordvik I., Grimholt U., Fosse V.M., Lie Y., Endresen C.;
RT "Cloning and sequence analysis of cDNAs encoding the MHC class II a-
chain in Atlantic salmon, Salmo salar.";
RL Immunogenetics 37:437-441(1993).
DR EMBL; X70166; CAA49725.1; -.
DR HSSP; P01888; 1BMG.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00969; MHC_II_beta; 1.
DR ProDom; PD000328; MHC_II_beta; 1.
```

```
DR SMART; SM00407; IGc1; 1.
KW Glycoprotein; MHC II; Transmembrane.
SQ SEQUENCE 244 AA; 27449 MW; 496CB9EA9D73765C CRC64;

Query Match 60.1%; Score 44.5; DB 7; Length 244;
Best Local Similarity 55.6%; Pred. No. 6.4;
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 1 QYIKANS---KFIGITEL 15
: ||: || ||: ||||
51 EYIRFNSTVGKFGVGYTEL 68

Db          :||: || ||: ||||
51 EYIRFNSTVGKFGVGYTEL 68

RESULT 8
Q9XG37      PRELIMINARY;      PRT;      546 AA.
AC Q9XG37;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical 66.2 kDa protein.
OS Guillardia theta (Cryptomonas phi).
OC Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.
OX NCBI_TaxID=55529;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20087226; PubMed=10618395;
RA Zauner S., Fraunholz M., Wastl J., Penny S.L., Beaton M.,
RA Cavalier-Smith T., Maier U., Douglas S.;
RT "Chloroplast protein and centrosomal genes, a tRNA intron, and odd
RT telomeres in an unusually compact eukaryotic genome, the cryptomonad
RT nucleomorph.".
RL Proc. Natl. Acad. Sci. U.S.A. 97:200-205(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=2123349; PubMed=11323671;
RA Douglas S., Zauner S., Fraunholz M., Beaton M., Penny S., Deng L.,
RA Wu X., Reith M., Cavalier-Smith T., Maier U.;
RT "The highly reduced genome of an enslaved algal nucleus.";
RL Nature 410:1091-1096(2001).
DR EMBL; AJ010592; CAB40403.1; -.
KW Hypothetical protein.
SQ SEQUENCE 546 AA; 66218 MW; 7303950F632BE6F2 CRC64;

Query Match 59.5%; Score 44; DB 10; Length 546;
Best Local Similarity 50.0%; Pred. No. 18;
Matches 7; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 2 YIKANSKFIGITEL 15
: ||: || ||: ||||
445 FIKNSRFRMRLTEI 458

Db          :||: || ||: ||||
445 FIKNSRFRMRLTEI 458

RESULT 9
Q31578      PRELIMINARY;      PRT;      67 AA.
AC Q31578;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE (DB03) MHC class II beta 1 (Fragment).
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RA Grimholt U., Olsaker I., de Vries Lindstrom C., Lie O.;
RT "A study of polymorphism in the MHC class II beta 1 and MHC class I
RT alpha 2 domain exons of Atlantic salmon (Salmo salar).";
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; L24929; AAA49550.1; -.
```

DR InterPro; IPR000353; MHC II beta.
 DR Pfam; PF00969; MHC II beta; 1.
 DR ProDom; PD000328; MHC II beta; 1.
 KW Glycoprotein; MHC II; Transmembrane.
 FT NON_TER 1
 FT 67
 SQ SEQUENCE 67 AA; 7449 MW; 42771AEDBABA6626 CRC64;
 Query Match 58.1%; Score 43.5; DB 7; Length 67;
 Best Local Similarity 50.0%; Pred. No. 2.6; 2; Indels 3; Gaps 1;
 Matches 9; Conservative 4; Mismatches 2;
 QY 1 QYIKANS---KFIGITEL 15
 : : : : :
 Db 16 EYVENSTVGKFGVYTEL 33
 : : : : :
 RESULT 10
 Q9CF66 PRELIMINARY; PRT; 180 AA.
 ID Q9CF66
 AC Q9CF66
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Spermidine acetyltransferase (EC 2.3.1.57).
 GN YOFF OR LL1615.
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Streptococcaceae; Lactococcus.
 OX NCBI_TaxID=1360;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IL1403;
 RX MEDLINE=21235186; PubMed=1137471;
 RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malarne K.,
 RA Weissenbach J., Ehrlich S.D., Sorokin A.;
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus
 RT lactis ssp. lactis IL1403."
 RL Genome Res. 11:731-753(2001).
 DR ENBL; AE006391; AAK05713.1; -.
 DR InterPro; IPR000182; GCSacetyltransf.
 DR Pfam; PF00583; Acetyltransf; 1.
 KW Transferase; Complete proteome.
 SQ SEQUENCE 180 AA; 21022 MW; 6DBD148524CODEF3C CRC64;
 Query Match 58.1%; Score 43; DB 16; Length 180;
 Best Local Similarity 69.2%; Pred. No. 8.8;
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 3 IKANSKFIGITEL 15
 : : : : :
 Db 65 IEANDTFIGIVEL 77
 : : : : :
 RESULT 11
 Q9MCL7 PRELIMINARY; PRT; 250 AA.
 ID Q9MCL7
 AC Q9MCL7
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE ORF13.
 GN ORF13.
 OS Streptococcus thermophilus bacteriophage 7201.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
 OX NCBI_TaxID=112023;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20088830; PubMed=10620678;
 RA Stanley E., Walsh L., van der Zwet A., Fitzgerald G.F.,
 RA van Sinderen D.;
 RT "Identification of four loci isolated from two Streptococcus
 RT thermophilus phage genomes responsible for mediating bacteriophage

RT resistance.";
 RL PEMS Microbiol. Lett. 182:271-277(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Stanley E., Walsh L., Fitzgerald G.F., van Sinderen D.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 DR ENBL; AF145054; AAF43506.1; -.
 DR InterPro; IPR001091; CNA Metransf.
 DR InterPro; IPR002295; D21N6_mtfase.
 DR Pfam; PF01555; N6_N4_Mtase; 1.
 DR PRINTS; PRO0506; D21N6MTFRASE.
 DR PRINTS; PRO0508; S21N4MTFRASE.
 SQ SEQUENCE 250 AA; 28744 MW; 725EA3B2B56B0D7F CRC64;
 Query Match 58.1%; Score 43; DB 9; Length 250;
 Best Local Similarity 57.1%; Pred. No. 12;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 1 QYIKANSKFIGITE 14
 : : : : :
 Db 131 QVLKANKIVGATE 144
 : : : : :
 RESULT 12
 Q9XJE8 PRELIMINARY; PRT; 252 AA.
 ID Q9XJE8
 AC Q9XJE8
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Putative methylase.
 OS Lactococcus lactis bacteriophage Tuc2009.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
 OX NCBI_TaxID=35241;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA van Sinderen D., van de Guchte M., Seegers J.F.M.L., Fitzgerald G.F.;
 RT "Molecular analysis of the temperate lactococcal phage Tuc2009."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR ENBL; AF109874; AAD37103.1; -.
 DR InterPro; IPR001091; CNA Metransf.
 DR InterPro; IPR002295; D21N6_mtfase.
 DR Pfam; PF01555; N6_N4_Mtase; 1.
 DR PRINTS; PRO0506; D21N6MTFRASE.
 DR PRINTS; PRO0508; S21N4MTFRASE.
 KW Methyltransferase.
 SQ SEQUENCE 252 AA; 29357 MW; 4D5A59FF47363948 CRC64;
 Query Match 58.1%; Score 43; DB 9; Length 252;
 Best Local Similarity 57.1%; Pred. No. 12;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 1 QYIKANSKFIGITE 14
 : : : : :
 Db 130 QVLKANKIVGATE 143
 : : : : :
 RESULT 13
 Q9CRV4 PRELIMINARY; PRT; 302 AA.
 ID Q9CRV4
 AC Q9CRV4
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE 311003OA04Rik protein (Fragment).
 GN 311003OA04RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:05:43 ; Search time 0.395043 Seconds
(without alignments)
1574.881 Million cell updates/sec

Title: US-09-806-703a-12

Perfect score: 74

Sequence: 1 QYKANSKFIGITEL 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	74	100.0	1314	1	TETX_CLOTE
2	44	59.5	66	1	VG84_BPML5
3	42.5	57.4	1682	1	MSPI_PLAFL3
4	42.5	57.4	1701	1	MSPI_PLAFL3
5	42.5	57.4	1701	1	MSPI_PLAFL3
6	42.5	57.4	1726	1	MSPI_PLAFL3
7	42.5	57.4	1726	1	MSPI_PLAFL3
8	41	55.4	204	1	PYRC_SERMA
9	41	55.4	384	1	LE12_THETN
10	41	55.4	899	1	V120_HSVSA
11	40	54.1	194	1	ACPD_HABIN
12	40	54.1	601	1	PEFL_LACLC
13	40	54.1	601	1	PEPF_LACLC
14	40	54.1	644	1	YHJ9_YEAST
15	39	52.7	213	1	KAD_MYCA
16	38	51.4	256	1	YD83_METJA
17	38	51.4	287	1	TRUB_AQUAE
18	38	51.4	572	1	HEMA_P13HT
19	37	50.0	191	1	Y096_HABIN
20	37	50.0	445	1	GNT1_HUMAN
21	37	50.0	447	1	GNT1_MOUSE
22	37	50.0	447	1	GNT1_RAT
23	37	50.0	447	1	GNT1_RAT
24	37	50.0	490	1	Y032_BORBU
25	37	50.0	510	1	G6PD_ASPNG
26	37	50.0	511	1	G6PD_EMENI
27	37	50.0	548	1	YDD2_SCHPO
28	36	48.6	169	1	Y358_BUCAI
29	36	48.6	258	1	MIP_CHLPN
30	36	48.6	296	1	YD01_CLOAB
31	36	48.6	333	1	DPOB_XENLA
32	36	48.6	451	1	MURD_BACSU
33	36	48.6	461	1	NIFN_RHOCA

RESULT 1

ID	TETX_CLOTE	STANDARD;	PRT;	1314 AA.
AC	P04958;			
DT	13-AUG-1987 (Rel. 05, Created)			
DT	13-AUG-1987 (Rel. 05, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	Tetanus toxin precursor (EC 3.4.24.68) (Tentoxylisin).			
OS	Clostridium tetani.			
OG	Plasmid.			
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;			
OC	Clostridium			
OX	NCBI_TaxID=1513;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=87053814; PubMed=3536478;			
RA	Eisel U., Jarausch W., Goretzki K., Henschen A., Engels J.,			
RA	Weller U., Hudel M., Habermann E., Niemann H.;			
RT	"Tetanus toxin: primary structure, expression in E. coli, and			
RT	homology with botulinum toxins."			
RL	EMBO J. 5:2495-2502(1986).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CN3911;			
RX	MEDLINE=87040747; PubMed=3774547;			
RA	Fairweather N.F., Lyness V.A.;			
RT	"The complete nucleotide sequence of tetanus toxin."			
RL	Nucleic Acids Res. 14:7809-7812(1986).			
RN	[3]			
RP	SEQUENCE OF 742-1314 FROM N.A.			
RX	MEDLINE=86085672; PubMed=3510187;			
RA	Fairweather N.F., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.;			
RT	"Cloning, nucleotide sequencing, and expression of tetanus toxin			
RT	fragment C in Escherichia coli."			
RL	J. Bacteriol. 165:21-27(1986).			
RN	[4]			
RP	PARTIAL SEQUENCE, AND DISULFIDE BONDS.			
RX	MEDLINE=90201034; PubMed=2108021;			
RA	Kriegelstein K., Henschen A., Weller U., Habermann E.;			
RT	"Arrangement of disulfide bridges and positions of sulphydryl groups			
RT	in tetanus toxin."			
RL	Eur. J. Biochem. 188:39-45(1990).			
RN	[5]			
RP	PARTIAL SEQUENCE.			
RX	MEDLINE=92037649; PubMed=1935979;			
RA	Kriegelstein K.G., Henschen A.H., Weller U., Habermann E.;			
RT	"Limited proteolysis of tetanus toxin. Relation to activity and			
RT	identification of cleavage sites."			
RL	Eur. J. Biochem. 202:41-51(1991).			
RN	[6]			
RP	IDENTIFICATION AS ZINC-PROTEASE.			
RX	MEDLINE=93010948; PubMed=1396558;			
RA	Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,			
RA	Montecucco C.;			
RT	"Tetanus toxin is a zinc protein and its inhibition of			
RT	neurotransmitter release and protease activity depend on zinc."			

ALIGNMENTS

34	36	48.6	495	1	G6PD_PICJA
35	36	48.6	630	1	YND1_YEAST
36	36	48.6	774	1	RRP3_INCBE
37	36	48.6	774	1	RRP3_INCJJ
38	36	48.6	1630	1	MSPI_PLAFK
39	36	48.6	1639	1	MSPI_PLAFW
40	35	47.3	176	1	NU6C_SPIOL
41	35	47.3	200	1	ACD2_CLOAB
42	35	47.3	200	1	ACPD_SALTY
43	35	47.3	261	1	CABV_CHICK
44	35	47.3	294	1	CDD_ECOLI
45	35	47.3	321	1	YDG7_SCHPO

P11410	pichia jadi
P40009	saccharomyc
P21770	influenza c
P13877	influenza c
P04932	plasmodium
P04933	plasmodium
Q9m318	spinacia ol
Q97d01	clostridium
Q8xf04	salmonella
P04354	gallus gall
P13652	escherichia
Q10494	schizosacch

EMBO J. 11:3577-3583(1992).
[7] IDENTIFICATION OF SUBSTRATE.
MEDLINE=9063293; PubMed=1331807;
Schiavo G., Benfenati F., Poulin B., Rossetto O., de Laureto P.P.,
Rao Daegupta B.R., Montecucco C.,
RA "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT by proteolytic cleavage of synaptobrevin.";
RL Nature 359:832-835(1992).
[8]
X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
RP MEDLINE=97475217; PubMed=9334741;
RX Umland T.C., Wingert L.M., Swaminathan S., Furey W.F., Schmidt J.J.,
RA Sax M.,
RT "Structure of the receptor binding fragment HC of tetanus
RL neurotoxin.";
RN Nat. Struct. Biol. 4:788-792(1997).
CC -!- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
CC BOND OF SYNAPTOSOMAL-2.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF 76-GLN-|-PHE-77 BOND IN
CC SYNAPTOSOMAL-2.
CC -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
CC YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
CC AND ARE NON-TOXIC AFTER SEPARATION.
CC -!- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
CC GANGLIOSIDE RECEPTORS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; X04436; CAA28033.1; -;
DR EMBL; M12739; AAA23282.1; -;
DR EMBL; X06214; CAA29564.1; -;
DR PIR; A25689; BTCLTN
DR PDB; 1AF9; 29-APR-98.
DR PDB; 1A8D; 14-OCT-98.
DR MEROPS; M27.001; -;
DR InterPro; IPR000395; Bontoxilysin.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR PRODOM; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;
KW 3D-structure.
FT INIT.MET 0 0
FT CHAIN 1 456
FT CHAIN 457 1314
FT METAL 232 232
FT ACT_SITE 233 233
FT METAL 236 236
FT TRANSMEM 226 246
FT TRANSMEM 669 689
FT DISULFID 438 466
FT DISULFID 1076 1092
SQ SEQUENCE 1314 AA; 150550 MW; 134C3657133EF81D CRC64;

Query Match 100.0%; Score 74; DB 1; Length 1314;
Best Local Similarity 100.0%; Pred. No. 1.7e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITEL 15

Db 829 QYIKANSKFIGITEL 843
|||||
RESULT 2
VG84.BPMLS
ID _VG84.BPMLS STANDARD; PRT; 66 AA.
AC Q05301;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Gene 84 protein (GP84).
GN 84.
OS Mycobacteriophage L5.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC L5-like viruses.
OX NCBI_TaxID=31757;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93211282; PubMed=8459766;
RA Hatfull G.F., Sarkis G.J.;
RT "DNA sequence, structure and gene expression of mycobacteriophage L5:
RL Mol. Microbiol. 7:395-405(1993).
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; Z18946; CAA79460.1; -;
DR PIR; S31029; S31029.
SQ SEQUENCE 66 AA; 7424 MW; 9C7104C7A4FA7A5 CRC64;

Query Match 59.5%; Score 44; DB 1; Length 66;
Best Local Similarity 57.1%; Pred. No. 0.27;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 YIKANSKFIGITEL 15
|||||
Db 50 YIKRNGKFGVTWEV 63
|||||
RESULT 3
MSP1.PLAF3
ID MSP1.PLAF3 STANDARD; PRT; 1682 AA.
AC P19598; Q25921;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMWGA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate ro-33 / Ghana).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5834;
RN [1]
RP SEQUENCE OF 1-1061 FROM N.A.
RX MEDLINE=88166657; PubMed=3327688;
RA Certa U., Rotmann D., Matile H., Reber-Liske R.;
RT "A naturally occurring gene encoding the major surface antigen
RL precursor p190 of Plasmodium falciparum lacks tripeptide repeats.";
EMBO J. 6:4137-4142(1987).
RN [2]
RP SEQUENCE OF 1032-1682 FROM N.A.
RX MEDLINE=95354793; PubMed=7628566;
RA Tolle R., Bujard H., Cooper J.A.;
RT "Plasmodium falciparum: variations within the C-terminal region of
RL merozoite surface antigen-1.";
Exp. Parasitol. 81:47-54(1995).

CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (Potential).
 CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
 CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

DR EMBL; M35727; AAA29715.1; -.
 DR EMBL; Y00087; CAA68280.1; -.
 DR EMBL; Z35326; CAA84555.1; -.
 DR PIR; S06286; S06286.
 DR InterPro: IPR000561; EGF-like.
 DR Pfam; PF00008; EGF; 1.
 DR Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.

FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1682 MEROZOITE SURFACE PROTEIN 1.
 FT TRANSMEM 1666 1682 MEMBRANE ANCHOR (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 462 462 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 528 528 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 599 599 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 785 785 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 881 881 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 901 901 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 947 947 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1178 1178 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1569 1569 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1682 AA; 192462 MW; C82A1E159948CAD6 CRC64;

Query Match 57.4%; Score 42.5; DB 1; Length 1682;
 Best Local Similarity 60.0%; Pred. No. 13;
 Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

QY 1 QYIKANSKFI-GITE 14
 ::|||::|:
 Db 983 QFVKSNSKVITGLTE 997

RESULT 4
 MSP1_PLAFF STANDARD; PRT; 1701 AA.
 AC P13819;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
 DE (PMMSA).
 DE MSP-1.
 OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8814299; PubMed=2449612;
 RA Peterson M.G., Coppel R.L., McIntyre-P., Langford C.J., Woodrow G.,
 RA Brown G.V., Anders R.F., Kemp D.J.;
 RT "Variation in the precursor to the major merozoite surface antigens
 RT of Plasmodium falciparum."
 RL Mol. Biochem. Parasitol. 27:291-302(1988).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (Potential).
 CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
 CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

DR EMBL; M19143; AAA29653.1; -.
 DR PIR; A54498; A54498.
 DR InterPro: IPR000561; EGF-like.
 DR Pfam; PF00008; EGF; 1.
 DR Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.

FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.
 FT CARBOHYD 110 110 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 536 536 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 607 607 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 802 802 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 899 899 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 919 919 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 991 991 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1701 AA; 193719 MW; 3920B75E73D38552 CRC64;

Query Match 57.4%; Score 42.5; DB 1; Length 1701;
 Best Local Similarity 60.0%; Pred. No. 13;
 Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

QY 1 QYIKANSKFI-GITE 14
 ::|||::|:
 Db 1001 QFVKSNSKVITGLTE 1015

RESULT 5
 MSP1_PLAFF STANDARD; PRT; 1701 AA.
 AC P08569;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
 DE (PMMSA) (P190).
 DE MSP-1.
 OS Plasmodium falciparum (isolate mad20 / Papua New Guinea).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=70153;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88011243; PubMed=3079521;
 RA Tanabe K., Mackay M., Goman M., Scaife J.G.;
 RT "Allelic dimorphism in a surface antigen gene of the malaria parasite
 RT Plasmodium falciparum."
 RL J. Mol. Biol. 195:273-287(1987).
 RN [2]
 RP REVISIONS TO 1403; 1569 AND 1629.
 RA Tanabe K.;
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-115 FROM N.A.
 RX MEDLINE=86136024; PubMed=3004972;
 RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
 RA Stunnenberg H., Bujard H.;
 RT "Polymorphism of the precursor for the major surface antigens of
 RT Plasmodium falciparum merozoites: studies at the genetic level."
 RL EMBO J. 4:3823-3829(1985).


```
CC -----
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M37213; AAA29611.1; -.
CC InterPro: IPR000561; EGF-like.
CC Pfam: PF00008; EGF; 1.
CC Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;
CC Transmembrane; GPI-anchor.
CC -----
CC SIGNAL 1 19 POTENTIAL.
CC CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1.
CC CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 567 567 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 638 638 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 827 827 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 944 944 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 990 990 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 1016 1016 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 1221 1221 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 1613 1613 N-LINKED (GLCNAC. .) (POTENTIAL).
CC SQ SEQUENCE 1726 AA; 196174 MW; 5B59CEBFA2F9A026 CRC64;
CC -----
CC Query Match 57.4%; Score 42.5; DB 1; Length 1726;
CC Best Local Similarity 60.0%; Pred. No. 13;
CC Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;
CC -----
CC QY 1 QYIKANSKFI-GITE 14
CC Db 1026 QFVKSNSKVITGLTE 1040
CC -----
CC RESULT 8
CC ID PYRC SERMA STANDARD; PRT; 204 AA.
CC AC Q9S3S1;
CC DT 16-OCT-2001 (Rel. 40, Created)
CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC DE Dihydroorotase (EC 3.5.2.3) (DHOase) (Fragment).
CC PYRC.
CC OS Serratia marcescens.
CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Serratia.
CC NCBI_TaxID=615;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=SM6;
CC RA Berkmen M., Benedik M.J.;
CC RT "Dini inhibits transcription of Serratia marcescens nuclease.";
CC Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC CC -!- CATALYTIC ACTIVITY: (S)-dihydroorotate + H(2)O = N-carbamoyl-L-
CC aspartate.
CC CC -!- COFACTOR: THIS ENZYME TIGHTLY BINDS ONE ZINC ATOM PER CHAIN WHICH
CC IS REQUIRED FOR THE CATALYTIC MECHANISM. IT ALSO BINDS WEAKLY TO
CC TWO OTHERS. ZINCS WHICH ARE NOT ESSENTIAL FOR ACTIVITY (BY
CC SIMILARITY).
CC CC -!- PATHWAY: Pyrimidine biosynthesis; third step.
CC CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC CC -!- SIMILARITY: BELONGS TO THE DHOASE FAMILY. SUBFAMILY 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AFI75466; AAD50307.1; ALT_INIT.
CC InterPro: IPR002195; Dihydroorotase.
CC DR PROSITE; PS00482; DIHYDROOROTASE 1; PARTIAL.
CC DR PROSITE; PS00483; DIHYDROOROTASE 2; 1.
CC KW Pyrimidine biosynthesis; Hydrolase; Zinc.
CC FT NON TER 1 1
CC SQ SEQUENCE 204 AA; 22899 MW; 9DB0B6C9B834B310 CRC64;
CC -----
CC Query Match 55.4%; Score 41; DB 1; Length 204;
CC Best Local Similarity 46.2%; Pred. No. 2.9;
CC Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
CC -----
CC QY 1 QYIKANSKFI-GIT 13
CC Db 42 QYVQAGNRFLGAT 54
CC -----
CC RESULT 9
CC LE12 THETN STANDARD; PRT; 384 AA.
CC AC Q8RCF9;
CC DT 15-JUN-2002 (Rel. 41, Created)
CC DT 15-JUN-2002 (Rel. 41, Last sequence update)
CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
CC DE 2-isopropylmalate synthase 2 (EC 4.1.3.12) (Alpha-isopropylmalate
CC synthase 2) (Alpha-IPM synthetase 2).
CC GN LEU2A OR TTE0472.
CC OS Thermoanaerobacter tengcongensis.
CC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
CC Thermoanaerobacteriaceae; Thermoanaerobacter.
CC NCBI_TaxID=119072;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=MB4 / JCM 11007;
CC RX MEDLINE-2192816; PubMed-11997336;
CC RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
CC Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
CC Tan H., Chen R., Wang J., Yu J., Yang H.;
CC RT "A complete sequence of T. tengcongensis genome.";
CC RL Genome Res. 12:689-700(2002).
CC CC -!- FUNCTION: Catalyzes the condensation of the acetyl group of
CC acetyl-CoA with 3-methyl-2-oxobutanoate (2-oxoisovalerate) to form
CC 3-carboxy-3-hydroxy-4-methylpentanoate (2-isopropylmalate).
CC CC -!- CATALYTIC ACTIVITY: 3-carboxy-3-hydroxy-4-methylpentanoate + CoA =
CC acetyl-CoA + 3-methyl-2-oxobutanoate + H(2)O.
CC CC -!- PATHWAY: Leucine biosynthesis; first step.
CC CC -!- SUBUNIT: Homotetramer (By similarity).
CC CC -!- SIMILARITY: BELONGS TO THE ALPHA-IPM SYNTHETASE / HOMOCITRATE
CC SYNTHASE FAMILY. LEUA 1 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE013018; AM23753.1; -.
CC PROSITE; PS00815; AIPM_HOMOCIT_SYNTH_1; 1.
CC DR PROSITE; PS00816; AIPM_HOMOCIT_SYNTH_2; 1.
CC KW Leucine biosynthesis; Lyase; Complete proteome.
CC SQ SEQUENCE 384 AA; 42404 MW; 095310F2C0B4A4DD CRC64;
CC -----
CC Query Match 55.4%; Score 41; DB 1; Length 384;
CC Best Local Similarity 54.5%; Pred. No. 5.5;
CC Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AFI75466; AAD50307.1; ALT_INIT.
CC InterPro: IPR002195; Dihydroorotase.
CC DR PROSITE; PS00482; DIHYDROOROTASE 1; PARTIAL.
CC DR PROSITE; PS00483; DIHYDROOROTASE 2; 1.
CC KW Pyrimidine biosynthesis; Hydrolase; Zinc.
CC FT NON TER 1 1
CC SQ SEQUENCE 204 AA; 22899 MW; 9DB0B6C9B834B310 CRC64;
CC -----
CC Query Match 55.4%; Score 41; DB 1; Length 204;
CC Best Local Similarity 46.2%; Pred. No. 2.9;
CC Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
CC -----
CC QY 1 QYIKANSKFI-GIT 13
CC Db 42 QYVQAGNRFLGAT 54
CC -----
CC RESULT 9
CC LE12 THETN STANDARD; PRT; 384 AA.
CC AC Q8RCF9;
CC DT 15-JUN-2002 (Rel. 41, Created)
CC DT 15-JUN-2002 (Rel. 41, Last sequence update)
CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
CC DE 2-isopropylmalate synthase 2 (EC 4.1.3.12) (Alpha-isopropylmalate
CC synthase 2) (Alpha-IPM synthetase 2).
CC GN LEU2A OR TTE0472.
CC OS Thermoanaerobacter tengcongensis.
CC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
CC Thermoanaerobacteriaceae; Thermoanaerobacter.
CC NCBI_TaxID=119072;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=MB4 / JCM 11007;
CC RX MEDLINE-2192816; PubMed-11997336;
CC RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
CC Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
CC Tan H., Chen R., Wang J., Yu J., Yang H.;
CC RT "A complete sequence of T. tengcongensis genome.";
CC RL Genome Res. 12:689-700(2002).
CC CC -!- FUNCTION: Catalyzes the condensation of the acetyl group of
CC acetyl-CoA with 3-methyl-2-oxobutanoate (2-oxoisovalerate) to form
CC 3-carboxy-3-hydroxy-4-methylpentanoate (2-isopropylmalate).
CC CC -!- CATALYTIC ACTIVITY: 3-carboxy-3-hydroxy-4-methylpentanoate + CoA =
CC acetyl-CoA + 3-methyl-2-oxobutanoate + H(2)O.
CC CC -!- PATHWAY: Leucine biosynthesis; first step.
CC CC -!- SUBUNIT: Homotetramer (By similarity).
CC CC -!- SIMILARITY: BELONGS TO THE ALPHA-IPM SYNTHETASE / HOMOCITRATE
CC SYNTHASE FAMILY. LEUA 1 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE013018; AM23753.1; -.
CC PROSITE; PS00815; AIPM_HOMOCIT_SYNTH_1; 1.
CC DR PROSITE; PS00816; AIPM_HOMOCIT_SYNTH_2; 1.
CC KW Leucine biosynthesis; Lyase; Complete proteome.
CC SQ SEQUENCE 384 AA; 42404 MW; 095310F2C0B4A4DD CRC64;
CC -----
CC Query Match 55.4%; Score 41; DB 1; Length 384;
CC Best Local Similarity 54.5%; Pred. No. 5.5;
CC Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
```

```

OY 3 IKANSKFIGIT 13
DB 214 VKAGAKFVGVT 224

RESULT 10
ID V120 HSVSA STANDARD; PRT; 899 AA.
AC Q01055;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Capsid assembly protein 63.
GN 63 OR EERRF1.
OS Herpesvirus saimiri (strain 11).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10383;
RN 1
RP SEQUENCE FROM N.A.
RX MEDLINE=92333688; PubMed=1321287;
RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
RA Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA Honess R.W.;
RA "Primary structure of the herpesvirus saimiri genome.";
RL J. Virol. 66:5047-5058(1992).
RN 2
RP SEQUENCE FROM N.A.
RX MEDLINE=92230228; PubMed=1314457;
RA Nicholas J., Cameron K.R., Coleman H., Newman C., Honess R.W.;
RA "Analysis of nucleotide sequence of the rightmost 43 kbp of
RA herpesvirus saimiri (HVS) L-DNA: General conservation of genetic
RA organization between HVS and Epstein-Barr virus.";
RL Virology 188:296-310(1992).
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL37,
CC HSV-1 23, EBV BOLFI, VZV 21, HVS-1 63, AND HCMV UL47.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X64346; CAA45686.1; -.
CC EMBL; M86409; AAA46139.1; -.
CC F1R; G36812; G36812.
CC Capsid assembly.
SQ SEQUENCE 899 AA; 103350 MW; F1429B3770A2885E CRC64;

Query Match 55.4%; Score 41; DB 1; Length 899;
Best Local Similarity 50.0%; Pred. No. 13;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 QYIKANSKFIGITE 14
DB 124 QYITSNATFTGLSE 137

RESULT 11
ACPD HAEN
ID ACPD HAEN STANDARD; PRT; 194 AA.
AC P43013;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable acyl carrier protein phosphodiesterase (EC 3.1.4.14) (ACP
DE phosphodiesterase).
GN ACPD:OR H11366.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.

OY 1 QYIKANSKFIGITEL 15
DB 147 QYWKSLGFIGITDV 161

RESULT 12
PEFI LACLC
ID PEFI LACLC STANDARD; PRT; 601 AA.
AC P54124; P94880;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Oligoendopeptidase F, plasmid (EC 3.4.24.-).
GN PEPF1 OR PEPF.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG Plasmid pLP763.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1359;
RN 1
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX STRAIN=NCDO 763;
RX MEDLINE=95096044; PubMed=7798200;
RA Monnet V., Nardi M., Chopin A., Chopin M.-C., Gripon J.-C.;
RA "Biochemical and genetic characterization of PefF, an oligopeptidase

```

from *Lactococcus lactis*.";
 J. Biol. Chem. 269:32070-32076 (1994).
 [2]
 SEQUENCE FROM N.A.
 STRAIN=NCDO 763;
 MEDLINE=97352670; PubMed=9209029;
 Nardi M., Renault P., Monnet V.;
 "Duplication of the pepF gene and shuffling of DNA fragments on the
 lactose plasmid of *Lactococcus lactis*.";
 J. Bacteriol. 179:4164-4171 (1997).
 CC -1- FUNCTION: HYDROLYZES PEPTIDES CONTAINING BETWEEN 7 AND 17 AMINO
 ACIDS WITH A RATHER WIDE SPECIFICITY.
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announcement/
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z32522; CAA83534.1; -;
 CC EMBL: X99798; CAA68133.1; -;
 CC MEROPS: M03.007; -;
 CC InterPro: IPR004438; pepF.
 CC InterPro: IPR001567; Peptidase M3.
 CC InterPro: IPR000130; Zn_MTPeptidase.
 CC Pfam: PF01432; Peptidase M3; 1.
 CC TIGRFAMs: TIGR00181; pepF; 1.
 CC PROSITE: PS00142; ZINC_PROTEASE; 1.
 CC Hydrolase: Metalloprotease; Zinc; FALSE NEG.
 CC METAL 387 388 BY SIMILARITY (BY SIMILARITY).
 CC ACT SITE 388 388 BY SIMILARITY (BY SIMILARITY).
 CC METAL 391 391 ZINC (CATALYTIC) (BY SIMILARITY).
 CC METAL 394 394 ZINC (CATALYTIC) (BY SIMILARITY).
 CC CONFLICT 518 518 F -> S (IN REF. 2).
 CC SEQUENCE 601 AA; 69674 MW; C8B5C519FFA2F787 CRC64;
 DR
 DR EMBL: Z32522; CAA83534.1; -;
 DR EMBL: X99798; CAA68133.1; -;
 DR MEROPS: M03.007; -;
 DR InterPro: IPR004438; pepF.
 DR InterPro: IPR001567; Peptidase M3.
 DR InterPro: IPR000130; Zn_MTPeptidase.
 DR Pfam: PF01432; Peptidase M3; 1.
 DR TIGRFAMs: TIGR00181; pepF; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; FALSE NEG.
 DR Hydrolase: Metalloprotease; Zinc; Plasmid.
 DR METAL 387 388 BY SIMILARITY (BY SIMILARITY).
 DR ACT SITE 388 388 BY SIMILARITY (BY SIMILARITY).
 DR METAL 391 391 ZINC (CATALYTIC) (BY SIMILARITY).
 DR METAL 394 394 ZINC (CATALYTIC) (BY SIMILARITY).
 DR CONFLICT 518 518 F -> S (IN REF. 2).
 DR SEQUENCE 601 AA; 69674 MW; C8B5C519FFA2F787 CRC64;
 KW Hydrolase: Metalloprotease; Zinc; Plasmid.
 FT ACT SITE 388 388 BY SIMILARITY (BY SIMILARITY).
 FT METAL 391 391 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 394 394 ZINC (CATALYTIC) (BY SIMILARITY).
 FT CONFLICT 518 518 F -> S (IN REF. 2).
 SQ SEQUENCE 601 AA; 69674 MW; C8B5C519FFA2F787 CRC64;
 Query Match 54.1%; Score 40; DB 1; Length 601;
 Best Local Similarity 46.7%; Pred. No. 13;
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 QYIKANSKFIGITEL 15
 Db 284 RYIELRKKILGITDL 298
 :||:|:|:|:
 RESULT 13
 PEPPF LACLA STANDARD; PRT; 601 AA.
 AC Q9CEV7;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Oligoendopeptidase F homolog (EC 3.4.24.-).
 GN PEPPF OR LI1727.
 OS *Lactococcus lactis* (subsp. *lactis*) (Streptococcus lactis).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
 OX NCBI_TaxID=1360;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=IL1403;
 MEDLINE=21235186; PubMed=11337471;
 RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malarne K.,
 RA Weissenbach J., Ehrlich S.D., Sorokin A.;
 RT "The complete genome sequence of the lactic acid bacterium *Lactococcus*
 RT *lactis* ssp. *lactis* IL1403.";
 RL Genome Res. 11:731-753 (2001).
 CC -1- FUNCTION: HYDROLYZES PEPTIDES CONTAINING BETWEEN 7 AND 17 AMINO
 ACIDS WITH A RATHER WIDE SPECIFICITY (BY SIMILARITY).
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announcement/
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AE006403; AAK05825.1; -;
 CC InterPro: IPR004438; pepF.
 CC InterPro: IPR001567; Peptidase M3.
 CC InterPro: IPR000130; Zn_MTPeptidase.
 CC Pfam: PF01432; Peptidase M3; 1.
 CC TIGRFAMs: TIGR00181; pepF; 1.
 CC PROSITE: PS00142; ZINC_PROTEASE; 1.
 CC Hydrolase: Metalloprotease; Zinc; Complete proteome.
 CC METAL 387 388 ZINC (CATALYTIC) (BY SIMILARITY).
 CC ACT SITE 388 388 BY SIMILARITY (BY SIMILARITY).
 CC METAL 391 391 ZINC (CATALYTIC) (BY SIMILARITY).
 CC METAL 394 394 ZINC (CATALYTIC) (BY SIMILARITY).
 CC SEQUENCE 601 AA; 69883 MW; 045B6873F10FE19A CRC64;
 DR
 DR InterPro: IPR004438; pepF.
 DR InterPro: IPR001567; Peptidase M3.
 DR InterPro: IPR000130; Zn_MTPeptidase.
 DR Pfam: PF01432; Peptidase M3; 1.
 DR TIGRFAMs: TIGR00181; pepF; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 DR Hydrolase: Metalloprotease; Zinc; Complete proteome.
 DR METAL 387 388 ZINC (CATALYTIC) (BY SIMILARITY).
 DR ACT SITE 388 388 BY SIMILARITY (BY SIMILARITY).
 DR METAL 391 391 ZINC (CATALYTIC) (BY SIMILARITY).
 DR METAL 394 394 ZINC (CATALYTIC) (BY SIMILARITY).
 DR SEQUENCE 601 AA; 69883 MW; 045B6873F10FE19A CRC64;
 KW Hydrolase: Metalloprotease; Zinc; Complete proteome.
 FT ACT SITE 388 388 BY SIMILARITY (BY SIMILARITY).
 FT METAL 391 391 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 394 394 ZINC (CATALYTIC) (BY SIMILARITY).
 SQ SEQUENCE 601 AA; 69883 MW; 045B6873F10FE19A CRC64;
 Query Match 54.1%; Score 40; DB 1; Length 601;
 Best Local Similarity 46.7%; Pred. No. 13;
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 QYIKANSKFIGITEL 15
 Db 284 RYIELRKKILGITDL 298
 :||:|:|:|:
 RESULT 14
 YHJ9 YEAST STANDARD; PRT; 644 AA.
 AC P38694;
 DT 01-FEB-1995 (Rel. 3i, Created)
 DT 01-FEB-1995 (Rel. 3i, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical aldehyde-dehydrogenase like protein in FILL-VMA10
 DE intergenic region.
 GN YHR039C.
 OS *Saccharomyces cerevisiae* (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=S288C / AB972;
 MEDLINE=94378003; PubMed=8091229;
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
 RA Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,
 RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
 RA Latreille P., Louis E.J., Macri C., Mardis S., Mouser L.,
 RA Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
 RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
 RA Vaudin M.;
 RT "Complete nucleotide sequence of *Saccharomyces cerevisiae* chromosome
 RT VIII.";
 RL Science 265:2077-2082 (1994).
 CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announcement/
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U00062; AAB68915.1; -;
 DR
 DR PIR: S46746; S46746.

Search completed: July 22, 2003, 08:06:39
Job time : 2.39504 secs

```
DR SGD; S0001081; YHR039C.
DR InterPro; IPR002086; Aldenhyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Hypothetical protein; Oxidoreductase.
FT ACT SITE 354 354 BY SIMILARITY.
FT ACT SITE 389 389 BY SIMILARITY.
SQ SEQUENCE 644 AA; 71320 MW; 54DADDAEB2A16D4D CRC64;

Query Match 54.1%; Score 40; DB 1; Length 644;
Best Local Similarity 60.0%; Pred. No. 14;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 QYIKANSKFIGITEL 15
Db 38 QIIQNQKLIGITIL 52

RESULT 15
KAD_MYCCA STANDARD; PRT; 213 AA...
AC P10251;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Adenylate kinase (EC 2.7.4.3) (ATP-AMP transphosphorylase).
GN ADK.
OS Mycoplasma capricolum.
OC Bacteria; Firmicutes; Mollicutes; Entomoplasmatales;
OC Entomoplasmataceae.
OX NCBI_TaxID=2095;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 27343 / KID;
RA MEDLINE=88142549; PubMed=3481422;
RX Okubo S., Muto A., Kawauchi Y., Yamao F., Osawa S.;
RT "The ribosomal protein gene cluster of Mycoplasma capricolum.";
RL Mol. Gen. Genet. 210:314-322(1987).
CC -1- FUNCTION: THIS SMALL UBIQUITOUS ENZYME IS ESSENTIAL FOR
CC MAINTENANCE AND CELL GROWTH.
CC -1- CATALYTIC ACTIVITY: ATP + AMP = ADP + ADP.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE ADENYLATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X06414; CAA29724.1; -.
DR PIR; S02851; KIYMC.
DR HSP; P27142; 1ZIN.
DR InterPro; IPR000850; Adenylate_kin.
DR Pfam; PF00406; adenylatekinase; 1.
DR PRINTS; PR00094; ADENYLTKINASE.
DR PRODOM; PD000657; Adenylate_kin; 1.
DR PROSITE; PS00113; ADENYLATE_KINASE; 1.
KW Transferase; Kinase; ATP-binding.
FT NP BIND 7 15 ATP (BY SIMILARITY).
FT SEQUENCE 213 AA; 24616 MW; FCFPA426B6F92E16 CRC64;

Query Match 52.7%; Score 39; DB 1; Length 213;
Best Local Similarity 72.7%; Pred. No. 7;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy .2 YIKANSKFIGI 12
Db 185 YFTNSKPFIEI 193
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:05:43 ; Search time 0.55306 Seconds
(without alignments)
1574.881 Million cell updates/sec

Title: US-09-806-703A-14.

Perfect score: 112

Sequence: 1 FNNFTVSWLRVPKVSASHLE 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	112	100.0	1314	1	TETX_CLOTE
2	61	54.5	1274	1	BXF_CLOBO
3	61	54.5	1296	1	BXG_CLOBO
4	59	52.7	1295	1	BXA1_CLOBO
5	58	51.8	1290	1	BXB_CLOBO
6	57	50.9	1051	1	VP2_AHSV6
7	56	50.0	1250	1	BXE_CLOBO
8	56	50.0	1250	1	BXE_CLOBO
9	56	50.0	1295	1	BXA2_CLOBO
10	52	46.4	449	1	VNSS_INSVN
11	52	46.4	464	1	VNSS_TSWV1
12	52	46.4	467	1	VNSS_TSWVL
13	51	45.5	1196	1	BXCX_CLOBO
14	48	42.9	1290	1	BXC1_CLOBO
15	47.5	42.4	1276	1	BXD_CLOBO
16	47	42.0	760	1	AMY_CLOAB
17	46	41.1	1162	1	BXEN_CLOBO
18	46	41.1	1162	1	BXEN_CLOBO
19	45	40.2	639	1	CGMA_RHIME
20	44	39.3	241	1	CRTA_RHOCA
21	43	38.4	152	1	YOBID_ECOLI
22	43	38.4	867	1	RRPO_BYDV1
23	42	37.5	208	1	FTSQ_STRGR
24	42	37.5	789	1	A12M_YEAST
25	42	37.5	1337	1	YDMS_SCHPO
26	42	37.5	1539	1	SMCX_HUMAN
27	42	37.5	1548	1	SMCY_MOUSE
28	42	37.5	1554	1	SMCX_MOUSE
29	42	37.5	1560	1	SMCX_HUMAN
30	41.5	37.1	237	1	YBSO_YEAST
31	41.5	37.1	728	1	YJ88_YEAST
32	41	36.6	352	1	CITC_ECOLI
33	40.5	36.2	230	1	FIMC_SALTY

RESULT 1

ID	TETX_CLOTE	STANDARD;	PRT;	1314 AA.
AC	P04958;			
DT	13-AUG-1987 (Rel. 05, Created)			
DT	13-AUG-1987 (Rel. 05, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	Tetanus toxin precursor (EC 3.4.24.68) (Tentoxylisin).			
OS	Clostridium tetani.			
OG	Plasmid.			
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;			
OC	Clostridium.			
OX	NCBI_TaxID=1513;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=87053814; PubMed=3536478;			
RA	Eisel U., Jarausch W., Goretzki K., Henschen A., Engels J.,			
RA	Weller U., Hudel M., Habermann E., Niemann H.;			
RT	"Tetanus toxin: primary structure, expression in E. coli, and			
RT	homology with botulinum toxins.";			
RL	EMBO J. 5:2495-2502(1986).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-CN3911;			
RX	MEDLINE=87040747; PubMed=3774547;			
RA	Fairweather N.P., Lyness V.A.;			
RT	"The complete nucleotide sequence of tetanus toxin.";			
RL	Nucleic Acids Res. 14:7809-7812(1986).			
RN	[3]			
RP	SEQUENCE OF 742-1314 FROM N.A.			
RX	MEDLINE=86085672; PubMed=3510187;			
RA	Fairweather N.P., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.;			
RT	"Cloning, nucleotide sequencing, and expression of tetanus toxin			
RT	fragment C in Escherichia coli.";			
RL	J. Bacteriol. 165:21-27(1986).			
RN	[4]			
RP	PARTIAL SEQUENCE, AND DISULFIDE BONDS.			
RX	MEDLINE=90201034; PubMed=2108021;			
RA	Kriegelstein K., Henschen A., Weller U., Habermann E.;			
RT	"Arrangement of disulfide bridges and positions of sulphhydryl groups			
RT	in tetanus toxin.";			
RL	Eur. J. Biochem. 188:39-45(1990).			
RN	[5]			
RP	PARTIAL SEQUENCE.			
RX	MEDLINE=92037649; PubMed=1935979;			
RA	Kriegelstein K.G., Henschen A.H., Weller U., Habermann E.;			
RT	"Limited proteolysis of tetanus toxin. Relation to activity and			
RT	identification of cleavage sites.";			
RL	Eur. J. Biochem. 202:41-51(1991).			
RN	[6]			
RP	IDENTIFICATION AS ZINC-PROTEASE.			
RX	MEDLINE=93010948; PubMed=1396558;			
RA	Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,			
RA	Montecucco C.;			
RT	"Tetanus toxin is a zinc protein and its inhibition of			
RT	neurotransmitter release and protease activity depend on zinc.";			

Q09786 schizosacch
P92817 paralicthy
Q04276 saccharomyc
P47706 mycoplasma
O74456 schizosacch
P58952 drosophila
P37604 salmonella
P47049 saccharomyc
Q8tW29 methanopyru
P03470 influenza a
P03468 influenza a
P06819 influenza a

34 40.5 36.2 535 1 YA98 SCHPO
35 40 35.7 116 1 NU3M PAROL
36 40 35.7 118 1 YMX2 YEAST
37 40 35.7 269 1 PARA MYCGE
38 40 35.7 288 1 PEP1 SCHPO
39 40 35.7 383 1 G22C DROME
40 40 35.7 390 1 DACD SALTY
41 40 35.7 396 1 YJEB YEAST
42 40 35.7 402 1 LE22 METKA
43 40 35.7 453 1 NRAM IAWIL
44 40 35.7 454 1 NRAM IAPUE
45 40 35.7 469 1 NRAM IAPAR

ALIGNMENTS

RL EMBO J. 11:3577-3583(1992).
RN [7]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=93063293; PubMed=1331807;
RA Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA Dasgupta B.R., Montecucco C.;
RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT by proteolytic cleavage of synaptobrevin.";
RN Nature 359:832-835(1992).
RL [8]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
RX MEDLINE=97475217; PubMed=9334741;
RA Unland T.C., Wingert L.M., Swaminathan S., Furey W.F., Schmidt J.J.,
RA Sax M.;
RT "Structure of the receptor binding fragment HC of tetanus
RT neurotoxin.";
RL Nat. Struct. Biol. 4:788-792(1997).
CC -1- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
CC BOND OF SYNAPTOSOMAL-2.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 76-GLN-|-PHE-77 BOND IN
CC SYNAPTOSOMAL-2.
CC -1- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
CC YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
CC AND ARE NON-TOXIC AFTER SEPARATION.
CC -1- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
CC GANGLIOSIDE RECEPTORS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X04436; CAA28033.1; -
DR EMBL; M12739; AAA23282.1; -
DR EMBL; X06214; CAA29564.1; -
DR PIR; A25689; BTCLTN
DR PDB; 1AF9; 29-APR-98.
DR PDB; 1ABD; 14-OCT-98.
DR MEROPS; M27.001; -
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR000130; Zn.Mtpeptidase.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOKILYSIN.
DR PRODOM; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC PROTEASE; 1.
KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;
KW 3D-structure.
FT INIT MET 0 0
FT CHAIN 0 456
FT CHAIN 1 456 TETANUS TOXIN LIGHT CHAIN.
FT CHAIN 457 1314 TETANUS TOXIN HEAVY CHAIN.
FT METAL 232 232 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT SITE 233 233 BY SIMILARITY.
FT METAL 236 236 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 236 236 ZINC (CATALYTIC) (BY SIMILARITY).
FT TRANSMEM 226 246 POTENTIAL.
FT TRANSMEM 669 689 POTENTIAL.
FT DISULFID 438 466 INTERCHAIN.
FT DISULFID 1076 1092
SQ SEQUENCE 1314 AA; 150550 MW; 134C3657133EP81D CRC64;

Query Match 100.0%; Score 112; DB 1; Length 1314;
Best Local Similarity 100.0%; Pred. No. 1.8e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNNFTVSFMLRVPKVSASHLE 21

Db 946 FNNFTVSFMLRVPKVSASHLE 966
|||||
RESULT 2
BXF CLOBO
ID _BXF CLOBO STANDARD; PRT; 1274 AA.
AC P30596;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BoNT/F)
DE (Bontoxilysin F).
GN BOTF.
OS Clostridium botulinum.
OS Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 23387;
RX MEDLINE=93012902; PubMed=1398040;
RA East A.K., Richardson P.T., Allaway D., Collins M.D.,
RA Roberts T.A., Thompson D.E.;
RT "Sequence of the gene encoding type F neurotoxin of Clostridium
RT botulinum";
RL PEMS Microbiol. Lett. 75:225-230(1992).
RN [2]
RP SEQUENCE OF 1-64 FROM N.A.
RC STRAIN=Hobbs FT10;
RX MEDLINE=94297488; PubMed=7764998;
RA East A.K., Collins M.D.;
RT "Conserved structure of genes encoding components of botulinum
RT neurotoxin complex M and the sequence of the gene coding for the
RT nontoxic component in nonproteolytic Clostridium botulinum type F";
RN Curr. Microbiol. 29:69-77(1994).
RP SEQUENCE OF 634-1002 FROM N.A.
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulinum neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F";
RL J. Clin. Microbiol. 31:2255-2262(1993).
RN [4]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94230352; PubMed=8175689;
RA Yamaeaki S., Baumeister A., Binz T., Blas J., Link E., Cornille F.,
RA Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT F botulinum neurotoxins and tetanus toxin";
RL J. Biol. Chem. 269:12764-12772(1994).
CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-|-LYS-59
CC BOND OF SYNAPTOSOMAL-1 AND -2.
CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC detected action on small molecule substrates.
CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC BOTULINUM NEUROTOXIN; TYPES A, B, C1, D, E, F, AND G.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL; M92906; AAA23263.1; -;
 CC EMBL; S73676; AAC60475.1; -;
 CC EMBL; X70820; CAA50151.1; -;
 CC EMBL; X70816; CAA50147.1; -;
 CC HSP; P10845; 3BTA.
 CC MEROPS; M27.002; -;
 CC InterPro; IPR000395; Bontoxilysin.
 CC InterPro; IPR000130; Zn.MTpeptdase.
 CC Pfam; PF01742; Peptidase_M27; 1.
 CC PRINTS; PR00760; BONTOXILYSIN.
 CC ProDom; PD001963; Bontoxilysin; 1.
 CC PROSITE; PS00142; ZINC_PROTEASE; 1.
 CC Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
 CC CHAIN 1 436 BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.
 CC FT CHAIN 437 1274 BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.
 CC FT METAL 227 227 ZINC (CATALYTIC) (BY SIMILARITY).
 CC FT ACT SITE 228 228 BY SIMILARITY.
 CC FT METAL 231 231 ZINC (CATALYTIC) (BY SIMILARITY).
 CC FT DISULFID 429 445 INTERCHAIN (PROBABLE).
 CC SQ SEQUENCE 1274 AA; 146709 MW; 5B99756A7438B921 CRC64;

Query Match 54.5%; Score 61; DB 1; Length 1274;
 Best Local Similarity 57.1%; Pred. No. 0.049;
 Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVPK 14

DB 930 YQNFSSFWVRIPK 943

RESULT 3

BXG_CLOBO STANDARD; PRT; 1296 AA.
 AC Q60393;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Botulinum neurotoxin type G precursor (EC 3.4.24.69) (BONT/G)
 DE (Bontoxilysin G).
 GS BOTG.
 OS Clostridium botulinum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=113 / 30;
 RX MEDLINE=94092745; PubMed=8268233;
 RA Campbell K., Collins M.D., East A.K.;
 RT "Nucleotide sequence of the gene coding for Clostridium botulinum
 RT (Clostridium argentinense) type G neurotoxin: genealogical comparison
 RT with other clostridial neurotoxins.";
 RL Biochim. Biophys. Acta 1216:487-491(1993).
 CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
 CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
 CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
 CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
 CC ENDOPEPTIDASE.
 CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
 CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
 CC detected action on small molecule substrates.
 CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
 CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
 CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
 CC FORMATION AND TOXIN BINDING, RESPECTIVELY.
 CC -1- SUBCELLULAR LOCATION: Secreted (by similarity).
 CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF

CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; X74162; CAA52275.1; -;
 CC HSP; P10845; 3BTA.
 CC MEROPS; M27.002; -;
 CC InterPro; IPR000395; Bontoxilysin.
 CC InterPro; IPR000130; Zn.MTpeptdase.
 CC Pfam; PF01742; Peptidase_M27; 1.
 CC PRINTS; PR00760; BONTOXILYSIN.
 CC ProDom; PD001963; Bontoxilysin; 1.
 CC PROSITE; PS00142; ZINC_PROTEASE; 1.
 CC Neurotoxin; Hydrolase; Metalloprotease; Zinc.
 CC INIT MET 0 0 BY SIMILARITY.
 CC CHAIN 1 441 BOTULINUM NEUROTOXIN G, LIGHT-CHAIN.
 CC FT CHAIN 442 1296 BOTULINUM NEUROTOXIN G, HEAVY-CHAIN.
 CC FT METAL 229 229 ZINC (CATALYTIC) (BY SIMILARITY).
 CC FT ACT SITE 230 230 BY SIMILARITY.
 CC FT METAL 233 233 ZINC (CATALYTIC) (BY SIMILARITY).
 CC FT DISULFID 435 449 INTERCHAIN (PROBABLE).
 CC SQ SEQUENCE 1296 AA; 149013 MW; DC8E47E15F665C31 CRC64;

Query Match 54.5%; Score 61; DB 1; Length 1296;
 Best Local Similarity 38.1%; Pred. No. 0.049;
 Matches 8; Conservative 9; Mismatches 4; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVPKVSASHLE 21

DB 929 FDNFSINFWRTPKYNNDIQ 949

RESULT 4

BXAL_CLOBO STANDARD; PRT; 1295 AA.
 ID EXAL_CLOBO
 AC P10845; P18639; P01561;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BONT/A)
 DE (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
 DE chain; Botulinum neurotoxin A, heavy-chain].
 GN BOTA OR BNA OR ATX.
 OS Clostridium botulinum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCTC 2916;
 RX MEDLINE=90235864; PubMed=2185020;
 RA Thompson D.E., Brehm J.K., Oultram J.D., Swinfield T.-J.,
 RA Shone C.C., Atkinson T., Melling J., Minton N.P.;
 RT "The complete amino acid sequence of the Clostridium botulinum type A
 RT neurotoxin, deduced by nucleotide sequence analysis of the encoding
 RT gene.";
 RL Eur. J. Biochem. 189:73-81(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=62A;
 RX MEDLINE=90264400; PubMed=2160960;
 RA Binz B., Kuaronzo H., Wille M., Frevent J., Wernars K., Niemann H.;
 RT "The complete sequence of botulinum neurotoxin type A and comparison
 RT with other clostridial neurotoxins.";
 RL J. Biol. Chem. 265:9153-9158(1990).
 RN [3]

RP SEQUENCE OF 1-65 FROM N.A.
RC STRAIN=62A;
RX MEDLINE=97016817; PubMed=8863443;
RA East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT "Organization and phylogenetic interrelationships of genes encoding
RT components of the botulinum toxin complex in proteolytic Clostridium
RT botulinum types A, B, and F: evidence of chimeric sequences in the
RT gene encoding the nontoxic nonhemagglutinin component.";
RL Int. J. Syst. Bacteriol. 46:1105-1112(1996).
RN [4]
RP SEQUENCE OF 1-34 FROM N.A.
RC STRAIN=Hall;
RX MEDLINE=89350959; PubMed=2669749;
RA Betley M.J., Somers E., Dasgupta B.R.;
RT "Characterization of botulinum type A neurotoxin gene: delineation of
RT the N-terminal encoding region.";
RL Biochem. Biophys. Res. Commun. 162:1388-1395(1989).
RN [5]
RP SEQUENCE OF 1-18 FROM N.A.
RC STRAIN=Type A NIH;
RX MEDLINE=96096783; PubMed=8521962;
RA Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;
RT "Molecular characterization of two forms of nontoxic-nonhemagglutinin
RT components of Clostridium botulinum type A progenitor toxins.";
RL FEBS Lett. 376:41-44(1995).
RN [6]
RP SEQUENCE OF 1-16.
RX MEDLINE=84178501; PubMed=6370252;
RA Schmidt J.J., Sartymoorthy V., Dasgupta B.R.;
RT "Partial amino acid sequence of the heavy and light chains of
RT botulinum neurotoxin type A.";
RL Biochem. Biophys. Res. Commun. 119:900-904(1984).
RN [7]
RP SEQUENCE OF 1-46.
RA Dasgupta B.R., Foley J., Niece R.;
RT "Partial sequence of the light chain of botulinum neurotoxin type A.";
RL Biochemistry 26:4162-4162(1987).
RN [8]
RP SEQUENCE OF 1-5 AND 444-456.
RX MEDLINE=91120847; PubMed=2126206;
RA Dasgupta B.R., Dekleva M.B.;
RT "Botulinum neurotoxin type A: sequence of amino acids at the
RT N-terminus and around the nicking site.";
RL Biochimie 72:661-664(1990).
RN [9]
RP SEQUENCE OF 448-464 AND 872-895.
RX MEDLINE=8904662; PubMed=3178218;
RA Sartymoorthy V., Dasgupta B.R., Foley J., Niece R.L.;
RT "Botulinum neurotoxin type A: cleavage of the heavy chain into two
RT halves and their partial sequences.";
RL Arch. Biochem. Biophys. 266:142-151(1988).
RN [10]
RP SEQUENCE OF 448-482.
RX MEDLINE=85285016; PubMed=3896784;
RA Shone C.C., Hambleton P., Melling J.;
RT "Inactivation of Clostridium botulinum type A neurotoxin by trypsin
RT and purification of two tryptic fragments. Proteolytic action near
RT the COOH-terminus of the heavy subunit destroys toxin-binding
RT activity.";
RL Eur. J. Biochem. 151:75-82(1985).
RN [11]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94063091; PubMed=8243676;
RA Schiavo G., Santucci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA Benfenati F., Wilson M.C., Montecucco C.;
RT "Botulinum neurotoxins serotypes A and B cleave SNAP-25 at distinct
RT COOH-terminal peptide bonds.";
RL FEBS Lett. 335:99-103(1993).
RN [12]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94124495; PubMed=8294407;
RA Binz T., Blaszi J., Yanasaki S., Baumeister A., Link E., Suedhof T.C.,
RA Jahn R., Niemann H.;

RT "Proteolysis of SNAP-25 by types E and A botulin neurotoxins.";
RL J. Biol. Chem. 269:1617-1620(1994).
RN [13]
RP MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.
RX MEDLINE=21556941; PubMed=11700044;
RA Rigoni M., Caccin P., Johnson E.A., Montecucco C., Rossetto O.;
RT "Site-directed mutagenesis identifies active-site residues of the
RT light chain of botulinum neurotoxin type A.";
RL Biochem. Biophys. Res. Commun. 288:1231-1237(2001).
RN [14]
RP X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).
RX MEDLINE=98455071; PubMed=9783750;
RA Lacy D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;
RT "Crystal structure of botulinum neurotoxin type A and implications
RT for toxicity.";
RL Nat. Struct. Biol. 5:898-902(1998).
CC -1- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC binds with high affinity to peripheral neuronal presynaptic
CC membrane, is then internalized by receptor-mediated endocytosis.
CC The C-terminus of the heavy chain (H) is responsible for the
CC adherence of the toxin to the cell surface while the N-terminus
CC mediates transport of the light chain from the endocytic vesicle
CC to the cytosol. After translocation, the light chain (L)
CC hydrolyzes the 197-Gln-Arg-198 bond in SNAP-25, thereby blocking
CC neurotransmitter release. Inhibition of acetylcholine release
CC results in flaccid paralysis, with frequent heart or respiratory
CC failure.
CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC detected action on small molecule substrates.
CC -1- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC heavy chain (H).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- PHARMACEUTICAL: Available under the name BOTOX (Allergan) for
CC the treatment of strabismus and blepharospasm associated with
CC dystonia and cervical dystonia. Also used for the treatment of
CC hemifacial spasm and a number of other neurological disorders
CC characterized by abnormal muscle contraction.
CC -1- MISCELLANEOUS: There are seven antigenically distinct forms of
CC botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC -1- DATABASE: NAME=BOTOX product information web site;
CC WWW="http://www.botox.com/index.jsp?hp&productinfo".
CC -1- DATABASE: NAME=Protein Spotlight;
CC NOTE=Issue 19 of February 2002;
CC WWW="http://www.expasy.org/spotlight/articles/spotlt019.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52066; CA336289.1; -
DR EMBL; M30196; AAA23262.1; -
DR EMBL; X92973; CAA63551.1; -
DR EMBL; D67030; BAA1051.1; -
DR EMBL; M27892; AAA23269.1; -
DR PIR; A35294; BTCLAB.
DR PIR; S09492; S09492.
DR PDB; 3BTA; 01-OCT-99.
DR MEROPS; M27.002; -
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR000130; Zn_MTPeptidase.
DR Pfam; PF01742; Peptidase M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW Pharmaceutical; 3D-structure.
FT INIT_MET 0


```

FT CHAIN 1 447 BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT CHAIN 448 1295 BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT METAL 222 ZINC (CATALYTIC).
FT ACT_SITE 223
FT METAL 226 ZINC (CATALYTIC).
FT METAL 261 ZINC (CATALYTIC).
FT DISULFID 429 453 INTERCHAIN.
FT DISULFID 1234 1279
FT TRANSMEM 626 646
FT TRANSMEM 655 675
FT VARIANT 26 26
FT MUTAGEN 261 261
FT MUTAGEN 265 265 E-2A: DRASTIC DECREASE IN ENZYMIC ACTIVITY.
FT MUTAGEN 365 365 F-2A: DECREASE IN ENZYMIC ACTIVITY.
FT CONFLICT 1 1 Y-2A: DECREASE IN ENZYMIC ACTIVITY.
FT CONFLICT 479 479 P -> Q (IN REF. 1).
FT CONFLICT 875 875 E -> P (IN REF. 9).
FT CONFLICT 891 891 T -> L (IN REF. 8).
FT CONFLICT 891 891 S -> K (IN REF. 8).
SQ SEQUENCE 1295 AA; 149322 MW; 858342F754862579 CRC64;

Query Match 52.7% Score 59; DB 1; Length 1295;
Best Local Similarity 57.1% Pred. No. 0.11;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVVK 14
Db 937 YENFSTFWIRPK 950

RESULT 5
BXB CLOBO
ID -BXB CLOBO STANDARD; PRT; 1290 AA.
AC P10844; P10843;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type B precursor (EC 3.4.24.69) (BoNT/B)
DE (Bontoxilysin B).
GN BOTB.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=92384550; PubMed=1514783;
RA Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,
RA Whelan N.P.;
RT "Molecular cloning of the Clostridium botulinum structural gene
RT encoding the type B neurotoxin and determination of its entire
RT nucleotide sequence."
RL Appl. Environ. Microbiol. 58:2345-2354 (1992).
RN [2]
RN SEQUENCE OF 35-245 FROM N.A.
RC STRAIN=NTC 7273;
RA Szabo E.A., Pemberton J.M., Desmarchelier P.M.;
RL Submitted (APR-1992) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE OF 633-993 FROM N.A.
RC STRAIN=NTC 7273;
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulinum neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F."
RL J. Clin. Microbiol. 31:2255-2262 (1993).
RN [4]
RN SEQUENCE OF 1-44 AND 441-466.
RC STRAIN=657;
RX MEDLINE=89000987; PubMed=3139097;
RA Dasgupta B.R., Datta A.;
RT "Botulinum neurotoxin type B (strain 657): partial sequence and
RT similarity with tetanus toxin."

```

```

RL Biochimie 70:811-817 (1989).
RN [5]
RN SEQUENCE OF 1-16 AND 441-458.
RC STRAIN=OKRA;
RX MEDLINE=85197963; PubMed=3888113;
RA Schmidt J.J., Sathyanarayanan V., Dasgupta B.R.;
RT "Partial amino acid sequences of botulinum neurotoxins types B and
RT E."
RL Arch. Biochem. Biophys. 238:544-548 (1985).
RN [6]
RN IDENTIFICATION AS ZINC-PROTEASE.
RX MEDLINE=93054694; PubMed=1429690;
RA Schiavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;
RT "Botulinum neurotoxins are zinc proteases."
RL J. Biol. Chem. 267:23479-23483 (1992).
RN [7]
RN IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=93063293; PubMed=1331807;
RA Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA Dasgupta B.R., Montecucco C.;
RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT by proteolytic cleavage of synaptobrevin."
RL Nature 359:832-835 (1992).
CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-|-PHE-77 BOND OF
CC SYNAPTOSOMAL-2.
CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC detected action on small molecule substrates.
CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M81186; AAA23211.1; -
CC EMBL; Z11934; CAA77591.1; -
CC EMBL; X70817; CAA50148.1; -
CC PIR; S07128; S07128.
CC PIR; S07155; S07155.
CC PIR; S08562; S08562.
CC PIR; S08573; S08573.
CC PIR; S08574; S08574.
CC PIR; A48940; A48940.
CC HSSP; P10845; 3BTA.
CC MEROPS; M27.002; -.
CC InterPro; IPR000395; Bontoxilysin.
CC InterPro; IPR000130; Zn.MTpeptidase.
CC Pfam; PF01742; Peptidase_M27; 1.
CC PRINTS; PR00760; BONTOXILYSIN.
CC ProDom; PD001963; Bontoxilysin; 1.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
CC INIT MET 0 0
CC CHAIN 1 440 BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.
CC CHAIN 441 1290 BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.
CC METAL 229 229 ZINC (CATALYTIC) (BY SIMILARITY).
CC ACT_SITE 230 230 BY SIMILARITY.

```

```
FT METAL 233 233 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 436 445 INTERCHAIN (PROBABLE).
FT CONFLICT 29 29 T -> M (IN REF. 4).
FT CONFLICT 217 217 R -> G (IN REF. 2).
FT CONFLICT 224 224 A -> S (IN REF. 2).
FT CONFLICT 463 463 S -> R (IN REF. 4).
SQ SEQUENCE 1290 AA; 150670 MW; D21746E2C024DF43 CRC64;

Query Match 51.8%; Score 58; DB 1; Length 1290;
Best Local Similarity 64.3%; Pred.No. 0.15;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FNNFTVSFWLRVPK 14
Db 922 FLDVSFVIRIPK 935

RESULT 6
VP2_AHSV6
ID VP2_AHSV6 STANDARD; PRT; 1051 AA.
AC O71024;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Outer capsid protein VP2.
GN S2 OR L2.
OS African horse sickness virus 6 (AHSV-6) (African horse sickness virus
OS (serotype 6)).
OC Viruses; dsRNA viruses; Reoviridae; Orbivirus.
OX NCBI_TaxID=86060;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9827831; PubMed=9617769;
RA Williams C.F., Inoue T., Lucas A.-M., Zanotto P., Roy P.;
RT "The complete sequence of four major structural proteins of African
RT horse sickness virus serotype 6: evolutionary relationships within
RT and between the orbiviruses.";
RL Virus Res. 53:53-73(1998).
CC -1- FUNCTION: THE VP2 PROTEIN IS ONE OF THE TWO PROTEINS (WITH VP5)
CC WHICH CONSTITUTE THE VIRUS PARTICLE OUTER CAPSID. IT IS THE
CC MAJOR TARGET OF THE HOST IMMUNOGENIC RESPONSE.
CC -1- SIMILARITY: BELONGS TO THE REOVIRUSES VP2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; AF021235; AAC40994.1;
DR InterPro; IPR001742; Orbi_VP2.
DR Pfam; PF00898; Orbi_VP2; 1.
KW Coat protein.
SQ SEQUENCE 1051 AA; 122326 MW; 2B04DB9E389F4B5F CRC64;

Query Match 50.9%; Score 57; DB 1; Length 1051;
Best Local Similarity 47.6%; Pred.No. 0.18;
Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 1 FNNFTVSFWLRVPKVSASHLE 21
Db 636 FSKRFVSFWYRVEKITKHLE 656

RESULT 7
BXE_CLOBO
ID BXE_CLOBO STANDARD; PRT; 1250 AA.
AC Q00436;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
```

```
DE Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE (Bontoxilysin B).
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Beluga;
RX MEDLINE=92181428; PubMed=1543481;
RA Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT "Sequences of the botulinum neurotoxin E derived from Clostridium
RT botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT ATCC 43181 and ATCC 43755).";
RL Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92174922; PubMed=1541280;
RA Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;
RT "The complete amino acid sequence of the Clostridium botulinum type-E
RT neurotoxin, derived by nucleotide-sequence analysis of the encoding
RT gene.";
RL Eur. J. Biochem. 204:657-667(1992).
RN [3]
RP SEQUENCE OF 1-251 FROM N.A.
RX MEDLINE=90264400; PubMed=2160960;
RA Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;
RT "The complete sequence of botulinum neurotoxin type A and comparison
RT with other clostridial neurotoxins.";
RL J. Biol. Chem. 265:9153-9158(1990).
RN [4]
RP SEQUENCE OF 1-13.
RX MEDLINE=85197963; PubMed=3888113;
RA Schmidt J.J., Sathiyamoorthy V., Dasgupta B.R.;
RT "Partial amino acid sequences of botulinum neurotoxins types B and
RT E.";
RL Arch. Biochem. Biophys. 238:544-548(1985).
RN [5]
RP SEQUENCE OF 419-426.
RX MEDLINE=90344918; PubMed=2116911;
RA Gimenez J.A., Dasgupta B.R.;
RT "Botulinum neurotoxin type E fragmented with endoproteinase Lys-C
RT reveals the site trypsin nicks and homology with tetanus
RT neurotoxin.";
RL Biochimie 72:213-217(1990).
RN [6]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94063091; PubMed=8243676;
RA Schiavo G., Santucci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA Benfenati F., Wilson M.C., Montecucco C.;
RT "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT COOH-terminal peptide bonds.";
RL FEBS Lett. 335:99-103(1993).
RN [7]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94124495; PubMed=8294407;
RA Binz T., Biasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA Jahn R., Niemann H.;
RT "Proteolysis of SNAP-25 by types E and A botulin neurotoxins.";
RL J. Biol. Chem. 269:1617-1620(1994).
CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-|-ILE-
CC 181 BOND IN SNAP-25.
CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC detected action on small molecule substrates.
CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
```

FORMATION AND TOXIN BINDING, RESPECTIVELY.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF

CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X62089; CAA43999.1; -

DR EMBL; X62683; CAA44558.1; -

DR PIR; A60027; A60027.

DR PIR; B35294; B35294.

DR PIR; JH0257; JH0257.

DR PIR; S08575; S08575.

DR PIR; S18111; S18111.

DR PIR; S21178; S21178.

DR HSSP; P10845; 3BTA.

DR MSERP; M27.002; -

DR InterPro; IPR000395; Bontoxilysin.

DR InterPro; IPR000130; Zn_Mtpeptidse.

DR Pfam; PF01742; Peptidase M27; 1.

DR PRINTS; PR00760; Bontoxilysin.

DR ProDom; PD001963; Bontoxilysin; 1.

DR PROSITE; PS00142; ZINC_PROTEASE; 1.

KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.

FT INIT MET 0

FT CHAIN 1 421

FT CHAIN 422 1250

FT METAL 211 211

FT ACT_SITE 212 212

FT METAL 215 215

FT DISULFID 411 425

FT CONFLICT 176 176

FT CONFLICT 197 197

FT CONFLICT 339 339

FT CONFLICT 772 772

FT CONFLICT 962 962

FT CONFLICT 966 966

FT CONFLICT 1194 1194

SQ SEQUENCE 1250 AA; 143712 MW; D9FCE26DDA041EB4 CRC64;

Query Match 50.0%; Score 56; DB 1; Length 1250;

Best Local Similarity 53.8%; Pred. No. 0.32;

Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FNNFTVSFWLRVP 13

Db 911 YKNFSISFWIRIP 923

RESULT 8

BXE_CLOBU

ID_BXE_CLOBU STANDARD; PRT; 1250 AA.

AC P30995;

DT 01-JUL-1993 (Rel. 26, Created)

DT 15-JUN-2002 (Rel. 26, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BONT/E)

DE (Bontoxilysin E).

OS Clostridium butyricum.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.

OX NCBI_TaxID=1492;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 43181, and ATCC 43755;

RX MEDLINE=92181428; PubMed=1543481;

RA Poullet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;

RT "Sequences of the botulin neurotoxin E derived from Clostridium

RT botulinum type E (strain Beluga) and Clostridium butyricum (strains

RT ATCC 43181 and ATCC 43755).";

RL Biochem. Biophys. Res. Commun. 183:107-113(1992).

RN [2]

RP SEQUENCE OF 1-251 FROM N.A.

RC STRAIN=BL6340;

RX MEDLINE=91237316; PubMed=2033376;

RA Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,

RA Yokosawa N., Yashiki T., Oguma K.;

RT "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum

RT type E toxin gene from Clostridium butyricum strain BL6340.";

RL J. Gen. Microbiol. 137:519-525(1991).

RN [3]

RP SEQUENCE OF 1-48.

RC STRAIN=5262;

RA Gimenez J., Foley J., Dasgupta B.R.;

RT "Neurotoxin type E from Clostridium botulinum and C. butyricum;

RT partial sequence and comparison.";

RL FASEB J. 2:A1750-A1750(1988).

CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER

CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES. IS INTERNALIZED

CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD

CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT

CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC

CC ENDOPEPTIDASE.

CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the

CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No

CC detected action on small molecule substrates.

CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A

CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,

CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL

CC FORMATION AND TOXIN BINDING, RESPECTIVELY.

CC -1- SUBCELLULAR LOCATION: Secreted

CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF

CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X62088; CAA43998.1; -

DR EMBL; X53180; CAA37321.1; -

DR PIR; JH0256; JH0256.

DR PIR; S16145; S16145.

DR HSSP; P10845; 3BTA.

DR MSERP; M27.002; -

DR InterPro; IPR000395; Bontoxilysin.

DR InterPro; IPR000130; Zn_Mtpeptidse.

DR Pfam; PF01742; Peptidase M27; 1.

DR PRINTS; PR00760; BONTXILYSIN.

DR ProDom; PD001963; Bontoxilysin; 1.

DR PROSITE; PS00142; ZINC_PROTEASE; 1.

KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.

FT INIT MET 0

FT CHAIN 1 421

FT CHAIN 422 1250

FT METAL 211 211

FT ACT_SITE 212 212

FT METAL 215 215

FT DISULFID 411 425

FT CONFLICT 229 229

SQ SEQUENCE 1250 AA; 143265 MW; 8171B5B2C312857 CRC64;

Query Match 50.0%; Score 56; DB 1; Length 1250;

Best Local Similarity 53.8%; Pred. No. 0.32;

Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR000130; Zn M1peptidase.
DR Pfam; PF01742; Peptidase M27; 1.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; FALSE NEG.
KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT INIT_MET 0
FT CHAIN 1 447
FT CHAIN 448 1295
FT METAL 222 222
FT ACT_SITE 223 223
FT METAL 226 226
FT DISULFID 429 453
FT DISULFID 1234 1279
FT TRANSMEM 626 646
FT TRANSMEM 655 675
SQ SEQUENCE 1295 AA; 149279 MW; 5DA04A13D98D6372 CRC64;
Query Match 50.0%; Score 56; DB 1; Length 1295;
Best Local Similarity 50.0%; Pred. No. 0.33;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
QY 1 FNNFTVSFWLRVVK 14
DB 937 YENFTSFWKIPK 950
RESULT 10
VNSS_INSVN
ID_VNSS_INSVN STANDARD; PRT; 449 AA.
AC Q01811;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Nonstructural protein NS-5.
GN NSS.
OS Impatiens necrotic spot virus (strain NL-07) (INSV).
OC Viruses; ssRNA negative-strand viruses; Bunyaviridae; Tospovirus.
OX NCBI_TaxIDs=31622;
RN [1]_SEQUENCE FROM N.A.
RP MEDLINE=92331780; PubMed=1385787;
RX de Haan P., de Avila A.C., Kormelink R., Westerbroek A.,
RA Gielen J.J., Peters D., Goldbach R.;
RT "The nucleotide sequence of the S RNA of Impatiens necrotic spot
virus, a novel tospovirus".
RL FEBS Lett. 306:27-32(1992).
CC -!- FUNCTION: FORMS FILAMENTOUS INCLUSION BODIES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; X66972; CAA47382.1; -.
DR PIR; S23158; S23158.
DR InterPro; IPR004915; Bunya_NS-S_2.
DR Pfam; PF03231; Bunya_NS-S_2; 1.
KW Nonstructural protein.
SQ SEQUENCE 449 AA; 51197 MW; C46AC1372B114CA5 CRC64;
Query Match 46.4%; Score 52; DB 1; Length 449;
Best Local Similarity 45.0%; Pred. No. 0.49;
Matches 9; Conservative 5; Mismatches 4; Indels 2; Gaps 1;
QY 2 NN--FTVSFWLRVVKVSASH 19
DB 240 NNKPKFISLWIRPKIMKN 259

```

RESULT 11:
VNSS TSWV1 STANDARD; PRT; 464 AA.
AC P26002;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Nonstructural protein NS-S.
GN NS.
OS Tomato spotted wilt virus (strain Brazilian Br-01) (TSWV).
OC Viruses; ssRNA negative-strand viruses; Bunyaviridae; Tospovirus.
OX NCBI_TaxID=36413;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90264829; PubMed=1693160;
RA de Haan P., Wagmakers L., Peters D., Goldbach R.;
RT "The S RNA segment of tomato spotted wilt virus has an ambisense
character.";
RL J. Gen. Virol. 71:1001-1007(1990).
CC -!- FUNCTION: FORMS FILAMENTOUS INCLUSION BODIES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D00645; BAA00540.1; -.
DR PIR; JQ0547; MNVUWC.
DR InterPro; IPR004915; Bunya NS-S-2.
DR Pfam; PF03231; Bunya NS-S_2; 1.
KW Nonstructural protein.
SQ SEQUENCE 464 AA; 52448 MW; AE5519179F9BF377 CRC64;

Query Match 46.4%; Score 52; DB 1; Length 464;
Best Local Similarity 64.3%; Pred. No. 0.51;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 NNFTVSFWLRVPKV 15
Db 239 SHFKLSLWLRVPKV 252

RESULT 12
VNSS TSWV1 STANDARD; PRT; 467 AA.
AC P26003;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Nonstructural protein NS-S.
GN NS.
OS Tomato spotted wilt virus (strain Bulgarian L3) (TSWV).
OC Viruses; ssRNA negative-strand viruses; Bunyaviridae; Tospovirus.
OX NCBI_TaxID=36415;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91132150; PubMed=1993884;
RA Maiss E., Ivanova L., Breyel E., Adam G.;
RT "Cloning and sequencing of the S RNA from a Bulgarian isolate of
tomato spotted wilt virus.";
RL J. Gen. Virol. 72:461-464(1991).
CC -!- FUNCTION: FORMS FILAMENTOUS INCLUSION BODIES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D00645; BAA00540.1; -.
DR PIR; JQ0547; MNVUWC.
DR InterPro; IPR004915; Bunya NS-S-2.
DR Pfam; PF03231; Bunya NS-S_2; 1.
KW Nonstructural protein.
SQ SEQUENCE 464 AA; 52448 MW; AE5519179F9BF377 CRC64;

Query Match 46.4%; Score 52; DB 1; Length 464;
Best Local Similarity 64.3%; Pred. No. 0.51;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 NNFTVSFWLRVPKV 15
Db 239 SHFKLSLWLRVPKV 252

RESULT 13
BXCN_CLOBO STANDARD; PRT; 1196 AA.
ID BXCN_CLOBO STANDARD; PRT; 1196 AA.
AC P46081;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Botulinum neurotoxin type C1, nontoxic component.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Type C Stockholm / C-ST;
RX MEDLINE=92231894; PubMed=1567404;
RA Tsuzuki K., Kimura K., Fujii N., Yokosawa N., Oguma K.;
RT "The complete nucleotide sequence of the gene coding for the
RT nontoxic-nonhemagglutinin component of Clostridium botulinum type C
RT progenitor toxin.";
RL Biochem. Biophys. Res. Commun. 183:1273-1279(1992).
CC -!- FUNCTION: THE NONTXIC COMPONENT IS NECESSARY TO MAINTAIN
CC TOXICITY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X62389; CAA44262.1; -.
DR InterPro; IPR000395; Bontoxilysin.
DR Pfam; PF01742; Peptidase M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
KW Neurotoxin.
SQ SEQUENCE 1196 AA; 138740 MW; 4BD5956274D7F9C3 CRC64;

Query Match 45.5%; Score 51; DB 1; Length 1196;
Best Local Similarity 80.0%; Pred. No. 2.1;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 NNFTVSFWLR 11
Db 899 NNFTICFWLR 908

RESULT 14
BXCI_CLOBO STANDARD; PRT; 1290 AA.
ID BXCI_CLOBO STANDARD; PRT; 1290 AA.
AC P18640;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)

```


Search completed: July 22, 2003, 08:06:40
Job time : 1.55306 secs

RA MEDLINE=94230352; PubMed=8175689;
RA Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT F botulinum neurotoxins and tetanus toxin.";
RL J. Biol. Chem. 269:12764-12772(1994).
CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CLEAVES THE 60-LYS-|-LEU-61 BOND OF
CC SYNAPTOSOMES-1 AND -2.
CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neurocytotoxic apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC detected action on small molecule substrates.
CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC -1- MISCELLANEOUS: BOTULINUM TYPE D NEUROTOXIN IS SYNTHESIZED BY D
CC STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC BACTERIOPHAGE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X54254; CAA38175.1; -;
DR EMBL; S49407; AAB24244.1; -;
DR PIR; S11455; S11455.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -;
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR000130; Zn_Mtpeptidase.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT CHAIN 1 442 BOTULINUM NEUROTOXIN D, LIGHT-CHAIN.
FT METAL 443 1276 BOTULINUM NEUROTOXIN D, HEAVY-CHAIN.
FT ACT SITE 229 229 BY SIMILARITY.
FT ACT SITE 230 230 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 233 233 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 437 450 INTERCHAIN (PROBABLE).
FT VARIANT 15 16 ND -> PV (IN STRAIN D-SA).
FT VARIANT 17 18 ND -> LQ (IN STRAIN D-1873).
FT VARIANT 452 452 K -> Q (IN STRAIN D-SA).
FT VARIANT 457 457 R -> T (IN STRAIN D-SA).
FT VARIANT 457 457 R -> F (IN STRAIN D-1873).
FT VARIANT 462 462 A -> D (IN STRAIN D-1873).
FT VARIANT 489 489 K -> N (IN STRAIN CB16).
FT VARIANT 644 644 N -> K (IN STRAIN CB16).
FT VARIANT 1122 1122 Q -> R (IN STRAIN CB16).
SQ SEQUENCE 1276 AA; 146871 MW; C1EC50F46C8233E2 CRC64;
Query Match 42.4%; Score 47.5; DB 1; Length 1276;
Best Local Similarity 40.9%; Pred.No.8.3;
Matches 9; Conservative 7; Mismatches 5; Indels 1; Gaps 1;
QY 1 FNNFTVSFWLRVVK-VSASHLE 21
DB 928 YENSSVSFWIKSKDLTNSHNE 949

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:05:49 ; Search time 2.11464 Seconds
(without alignments)
2046.206 Million cell updates/sec

Title: US-09-806-703A-14
Perfect score: 112
Sequence: 1 FNNFTVSFWLRVPKVSASHLE 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPRTMBL 21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	112	100.0	451	2	Q9LA13
2	112	100.0	1310	2	Q93N27
3	62	55.4	1268	2	Q45851
4	61	54.5	1278	2	Q57236
5	58	51.8	361	2	Q45846
6	58	51.8	361	2	Q45848
7	58	51.8	441	2	Q9X708
8	58	51.8	1291	2	Q93G71
9	58	51.8	1291	2	Q933K0
10	58	51.8	1291	2	Q9ZAJ8
11	58	51.8	1291	2	Q08077
12	56	50.0	367	2	Q45861
13	56	50.0	367	2	Q45862
14	56	50.0	1251	2	Q9K395
15	56	50.0	1255	2	Q9FAR6
16	55	49.1	1280	2	Q9ZAJ5

17	52	46.4	467	12	Q37367	Q37367 tomato spot
18	52	46.4	467	12	Q37369	Q37369 tomato spot
19	51	45.5	467	12	Q88900	Q88900 tospovirus.
20	51	45.5	1196	2	Q9LBS8	Q9LBS8 clostridium
21	51	45.5	1196	2	Q9LBR2	Q9LBR2 clostridium
22	51	45.5	1196	2	Q53550	Q53550 clostridium
23	51	45.5	1196	2	Q45916	Q45916 clostridium
24	51	45.5	1196	2	Q93HT4	Q93HT4 clostridium
25	51	45.5	1196	9	Q9ZX77	Q9ZX77 clostridium
26	51	45.5	1196	9	Q38197	Q38197 clostridium
27	50	44.6	503	10	Q948E8	Q948E8 oryza sativ
28	49	43.8	276	5	Q97ZK5	Q97ZK5 caenorhabdi
29	48	42.9	504	10	Q9MIN3	Q9MIN3 arabidopsis
30	48	42.9	1285	2	Q9LBR1	Q9LBR1 clostridium
31	48	42.9	1285	2	Q45967	Q45967 clostridium
32	48	42.9	1291	2	Q93HT3	Q93HT3 clostridium
33	47.5	42.4	1275	12	Q9QTG7	Q9QTG7 clostridium
34	47.5	42.4	1280	2	Q9LBS7	Q9LBS7 clostridium
35	47.5	42.4	1280	2	Q45849	Q45849 clostridium
36	47	42.0	345	5	Q62165	Q62165 caenorhabdi
37	47	42.0	760	2	Q9S429	Q9S429 clostridium
38	46	41.1	278	2	Q93JMO	Q93JMO streptomyce
39	46	41.1	322	5	P91551	P91551 caenorhabdi
40	46	41.1	442	6	Q07452	Q07452 oviz aries
41	46	41.1	496	3	Q9UUI6	Q9UUI6 schizosacch
42	46	41.1	558	5	O76933	O76933 drosophila
43	46	41.1	558	5	Q9V3N9	Q9V3N9 drosophila
44	46	41.1	753	10	Q9C7W5	Q9C7W5 arabidopsis
45	46	41.1	996	10	Q9C925	Q9C925 arabidopsis

ALIGNMENTS

RESULT 1

Q9LA13 ID Q9LA13 PRELIMINARY; PRT; 451 AA.
AC Q9LA13;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Tetanus toxin (fragment).
OS Clostridium tetani.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=20886;
RA He H.J., Shi H.J., Fe Z.Y., Yuan Q.S., Wu X.F.;
RT "Fragment C of Tetanus Toxin."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF154828; AAF73267.1; -.
DR HSSP; P04958; IABD.
DR InterPro; IPR01064; Crystallin.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
FT NON TER 1
SQ SEQUENCE 451 AA; 51823 MW; 69A8C5F030B6CD8E CRC64;

Query Match 100.0%; Score 112; DB 2; Length 451;
Best Local Similarity 100.0%; Pred. No. 1.7e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FNNFTVSFWLRVPKVSASHLE 21
Db 83 FNNFTVSFWLRVPKVSASHLE 103

RESULT 2

Q93N27 ID Q93N27 PRELIMINARY; PRT; 1310 AA.
AC Q93N27;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Tetanus toxin (Fragment).
 OS Clostridium tetani.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
 OC Clostridiales; Clostridiaceae; Clostridium.
 OX NCBI_TaxID=1513;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Shumin Z., Dianliang L.;
 RT "Cloning and sequence analysis of tetanus toxin gene.";
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF389424; AAK72964.2; -
 DR InterPro; IPR000395; Bontoxilysin.
 DR InterPro; IPR001064; Crystallin.
 DR InterPro; IPR000130; Zn.MTpeptidase.
 DR Pfam; PF01742; Peptidase M27; 1.
 DR ProDom; PD001963; Bontoxilysin; 1.
 DR PROSITE; PS00225; CRYSTALLIN BETAGAMMA; UNKNOWN_1.
 DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
 FT NON_TER 1 1310
 FT NON_TER 1310 1310
 SQ SEQUENCE 1310 AA; 150316 MW; 9EADDC914418E450 CRC64;

Query Match 100.0%; Score 112; DB 2; Length 1310;
 Best Local Similarity 100.0%; Pred. No. 5.2e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVVKVSASHLE 21
 Db 948 FNNFTVSFWLRVVKVSASHLE 968

RESULT 3

Q45851 ID Q45851 PRELIMINARY; PRT; 1268 AA.
 AC Q45851;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Neurotoxin type F.
 GN BONT / F.
 OS Clostridium baratii.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
 OC Clostridiales; Clostridiaceae; Clostridium.
 OX NCBI_TaxID=1561;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9325228; PubMed=8486245;
 RA Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,
 RA Richardson P.T.;
 RT "Nucleotide sequence of the gene coding for Clostridium baratii type F
 neurotoxin: Comparison with other clostridial neurotoxins.";
 RL FEMS Microbiol. Lett. 108:175-182(1993).
 DR EMBL; X68262; CAA48329.1; -
 DR HSSP; P10845; 3BTA.
 DR MEROPS; M27.002; -
 DR InterPro; IPR000395; Bontoxilysin.
 DR InterPro; IPR000130; Zn.MTpeptidase.
 DR Pfam; PF01742; Peptidase M27; 1.
 DR PRINTS; PR00760; BONTXILYSIN.
 DR ProDom; PD001963; Bontoxilysin; 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
 SQ SEQUENCE 1268 AA; 145513 MW; 963040091AC15ED2 CRC64;

Query Match 55.4%; Score 62; DB 2; Length 1268;
 Best Local Similarity 64.3%; Pred. No. 0.13;
 Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVVK 14
 Db 922 YQNFSVFWVRPK 935

RESULT 4

Q57236 ID Q57236 PRELIMINARY; PRT; 1278 AA.
 AC Q57236; Q45863;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE Botulinum neurotoxin type F (BONT/F protein).
 GN BONT/F.
 OS Clostridium botulinum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
 OC Clostridiales; Clostridiaceae; Clostridium.
 OX NCBI_TaxID=1491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCTC 10281;
 RA Hutson R.A., Collins M.D.;
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Elmore M.J., Bodsworth N.J., Whelan S.M., Minton N.P.;
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 635-1000 FROM N.A.
 RC STRAIN=NCTC 1028;
 RX MEDLINE=94013172; PubMed=8408542;
 RA Campbell K., East A.K., Collins M.D.;
 RT "Gene probes for identification of the botulin neurotoxin gene and
 RT specific identification of neurotoxin types B, E, and F.";
 RL J. Clin. Microbiol. 31:2255-2262(1993).
 RN [4]
 RP SEQUENCE OF 1-27 FROM N.A.
 RC STRAIN=LANGEAND;
 RX MEDLINE=98404102; PubMed=9732534;
 RA East A.K., Bhandari M., Hielm S., Collins M.D.;
 RT "Analysis of the botulinum neurotoxin type F gene clusters in
 RT proteolytic and nonproteolytic Clostridium botulinum and Clostridium
 RT baratii.";
 RL Curr. Microbiol. 37:262-268(1998).
 DR EMBL; X81714; CAA57358.1; -
 DR EMBL; L35496; AAA23210.1; -
 DR EMBL; X70821; CAA50152.1; -
 DR EMBL; X99064; CAA67512.1; -
 DR HSSP; P10845; 3BTA.
 DR MEROPS; M27.002; -
 DR InterPro; IPR000395; Bontoxilysin.
 DR InterPro; IPR000130; Zn.MTpeptidase.
 DR Pfam; PF01742; Peptidase M27; 1.
 DR PRINTS; PR00760; BONTXILYSIN.
 DR ProDom; PD001963; Bontoxilysin; 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
 KW Neurotoxin.
 SQ SEQUENCE 1278 AA; 147073 MW; A1BE1318431D6918 CRC64;

Query Match 54.5%; Score 61; DB 2; Length 1278;
 Best Local Similarity 57.1%; Pred. No. 0.19;
 Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVVK 14
 Db 931 YQNFSVFWVRPK 944

RESULT 5

Q45846 ID Q45846 PRELIMINARY; PRT; 361 AA.
 AC Q45846;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE Botulinum neurotoxin type B (Fragment).
 GN BONT/B.

```

OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TYPE B;
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulin neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
DR ENBL; X70814; CAA50145.1; -.
DR HSSP; P10845; 3BTA.
KW Neurotoxin.
FT NON_TER 1 1
FT NON_TER 361 361
SQ SEQUENCE 361 AA; 42175 MW; 533EA98735CD98E1 CRC64;

Query Match 51.8%; Score 58; DB 2; Length 361;
Best Local Similarity 64.3%; Pred. No. 0.16;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVPK 14
Db 290 FLDFSVFWIRIPK 303

RESULT 6
Q45848 PRELIMINARY; PRT; 361 AA.
AC Q45848;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Botulinum neurotoxin type B (Fragment).
GN BONT/B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TYPE B;
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulin neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
DR ENBL; X70819; CAA50150.1; -.
DR HSSP; P10845; 3BTA.
KW Neurotoxin.
FT NON_TER 1 1
FT NON_TER 361 361
SQ SEQUENCE 361 AA; 42131 MW; A2E0FFFC81F9533D CRC64;

Query Match 51.8%; Score 58; DB 2; Length 361;
Best Local Similarity 64.3%; Pred. No. 0.16;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVPK 14
Db 290 FLDFSVFWIRIPK 303

RESULT 7
Q9X708 PRELIMINARY; PRT; 441 AA.
AC Q9X708;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Botulinum neurotoxin type B (Fragment).

```

```

GN BONT/B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=9343691; PubMed=10413679;
RA Lalli G., Herreros J., Osborne S.L., Montecucco C., Rossetto O.,
RA Schiavo G.;
RT "Functional characterisation of tetanus and botulinum neurotoxins
RT binding domains.";
RL J. Cell Sci. 112:2715-2724(1999).
DR EMBL; AJ242628; CAB43706.1; -.
DR HSSP; P10845; 3BTA.
KW Neurotoxin.
FT NON_TER 1 1
FT NON_TER 441 441
SQ SEQUENCE 441 AA; 52772 MW; 721D0B468E8C95A4 CRC64;

Query Match 51.8%; Score 58; DB 2; Length 441;
Best Local Similarity 64.3%; Pred. No. 0.2;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVPK 14
Db 81 FLDFSVFWIRIPK 94

RESULT 8
Q93G71 PRELIMINARY; PRT; 1291 AA.
AC Q93G71;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Neurotoxin type B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1436;
RA Kirma N., Ferreira J.L., Baumstark B.R.;
RT "Characterization of six type A strains of Clostridium botulinum that
RT contain type B toxin gene sequences.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR ENBL; AF295926; AAK97132.1; -.
DR InterPro; IPR000130; Zn.MTpeptdse.
DR Pfam; PF01742; Peptidase_M27; 1.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.
SQ SEQUENCE 1291 AA; 150824 MW; D7CA07BAE2EB8CD2 CRC64;

Query Match 51.8%; Score 58; DB 2; Length 1291;
Best Local Similarity 64.3%; Pred. No. 0.6;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVPK 14
Db 923 FLDFSVFWIRIPK 936

RESULT 9
Q933K0 PRELIMINARY; PRT; 1291 AA.
AC Q933K0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Type B cryptic neurotoxin.

```



```

RP SEQUENCE FROM N.A.
RC STRAIN=TYPE E;
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulinum neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
DR EMBL; X70818; CAA50149.1; -.
DR HSSP; P10845; 3BTA.
KW Neurotoxin.
FT NON_TER 1 1
FT NON_TER 367 367
SQ SEQUENCE 367 AA; 42902 MW; 346A610C2FF70262 CRC64;

Query Match 50.0%; Score 56; DB 2; Length 367;
Best Local Similarity 53.8%; Pred. No. 0.36;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVP 13
Db 297 YKNFSISFWVRIP 309

RESULT 13
Q45862 PRELIMINARY; PRT; 367 AA.
AC Q45862;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (T-EMBLrel. 15, Last annotation update)
DE Botulinum neurotoxin type E (Fragment).
GN BONT/E.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1491;
RN SEQUENCE FROM N.A.
RC STRAIN=TYPE E;
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulinum neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
DR EMBL; X70815; CAA50146.1; -.
DR HSSP; P10845; 3BTA.
KW Neurotoxin.
FT NON_TER 1 1
FT NON_TER 367 367
SQ SEQUENCE 367 AA; 42854 MW; 0810595B3A865570 CRC64;

Query Match 50.0%; Score 56; DB 2; Length 367;
Best Local Similarity 53.8%; Pred. No. 0.36;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVP 13
Db 297 YKNFSISFWVRIP 309

RESULT 14
Q9K395 PRELIMINARY; PRT; 1251 AA.
AC Q9K395;
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE Type E botulinum toxin.
GN BONT/E.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1492;

RN SEQUENCE FROM N.A.
RC STRAIN=LCL 095;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.;
RT "C. butyricum (LCL 095) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL 155;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Gyobu Y., Yamakawa K.,
RA Kato H., Nakamura S., Karasawa T.;
RT "C. butyricum (LCL 155) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL 155;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.;
RT "C. butyricum (LCL 155) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=KZ 1899;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.;
RT "C. butyricum (KZ 1899) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=KZ 1897;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.;
RT "C. butyricum (KZ 1897) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=KZ 1898;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.;
RT "C. butyricum (KZ 1898) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=KZ 1886;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.;
RT "C. butyricum (KZ 1886) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN=KZ 1887;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.;
RT "C. butyricum (KZ 1887) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE FROM N.A.
RC STRAIN=KZ 1889;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.;
RT "C. butyricum (KZ 1889) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [9]
RP SEQUENCE FROM N.A.
RC STRAIN=KZ 1890;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.;
RT "C. butyricum (KZ 1890) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [10]
RP SEQUENCE FROM N.A.
RC STRAIN=KZ 1891;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.;
RT "C. butyricum (KZ 1891) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [11]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL 063;

```

RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.,
RT "C. butyricum (ICL 063) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR ENBL; AB037714; BAB03522.1; -
DR ENBL; AB037704; BAB03512.1; -
DR ENBL; AB037705; BAB03513.1; -
DR ENBL; AB037706; BAB03514.1; -
DR ENBL; AB037707; BAB03515.1; -
DR ENBL; AB037708; BAB03516.1; -
DR ENBL; AB037709; BAB03517.1; -
DR ENBL; AB037710; BAB03518.1; -
DR ENBL; AB037711; BAB03519.1; -
DR ENBL; AB037712; BAB03520.1; -
DR ENBL; AB037713; BAB03521.1; -
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR000130; Zn_MTPeptidase.
DR Pfam; PF01742; Peptidase M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.
SQ SEQUENCE 1251 AA; 143751 MW; 2021F4E427070296 CRC64;

Query Match 50.0%; Score 56; DB 2; Length 1251;
Best Local Similarity 53.8%; Pred. No. 1.3;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVP 13
:|:|:|:|:|:
Db 912 YKNFSISFWVRIP 924

RESULT 15

Q9FAR6 PRELIMINARY; PRT; 1255 AA.
AC Q9FAR6;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE Type E botulinum toxin.
GN BONT/E.
OS Clostridium butyricum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1492;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BL 6340/ATCC 43755/BL 5520/KZ 147;
RX MEDLINE=20509829; PubMed=11055954;
RA Wang X., Maegawa T., Karasawa T., Kozaki S., Tsukamoto K., Gyobu Y.,
RA Yamakawa K., Oguma K., Sakaguchi Y., Nakamura S.,
RT "Genetic Analysis of Type E Botulinum Toxin-Producing Clostridium
RT butyricum Strains.";
RL Appl. Environ. Microbiol. 66:4992-4997(2000).
DR ENBL; AB039264; BAB12249.1; -
DR HSSP; P10845; 3BTA.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR000130; Zn_MTPeptidase.
DR Pfam; PF01742; Peptidase M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.
SQ SEQUENCE 1255 AA; 143918 MW; 1B557B9D85CD8E4D CRC64;

Query Match 50.0%; Score 56; DB 2; Length 1255;
Best Local Similarity 53.8%; Pred. No. 1.3;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVP 13
:|:|:|:|:|:
Db 915 YKNFSISFWVRIP 927

Search completed: July 22, 2003, 08:11:31
Job time : 4.11464 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:05:49 ; Search time 1.05732 Seconds
(without alignments)
1909.378 Million cell updates/sec

Title: US-09-806-703A-14
Perfect score: 112
Sequence: 1 FNNFTVSFWLRVPKVSASHLE 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73.*

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	112	100.0	1315	1 BTCLTN	tentoxilysin (EC 3
2	62	55.4	1268	2 S33411	botulinum neurotox
3	61	54.5	366	2 S48110	neurotoxin type F
4	61	54.5	369	2 S48109	neurotoxin type F
5	61	54.5	1274	2 I40813	neurotoxin - Clost
6	61	54.5	1297	2 S39791	neurotoxin - Clost
7	59	52.7	1296	1 BTCLAB	botulinum neurotox
8	58	51.8	1291	1 A48940	botulinum neurotox
9	58	51.8	1291	2 I40631	non-proteolytic bo
10	56	50.0	367	2 S48106	neurotoxin type E
11	56	50.0	1251	2 JH0256	botulinum neurotox
12	56	50.0	1252	2 S21178	botulinum neurotox
13	56	50.0	1296	2 I40645	botulinum neurotox
14	52	46.4	449	2 S23158	nucleocapsid prote
15	52	46.4	464	1 MNVUWC	nonstructural prot
16	52	46.4	467	1 MNVUW1	nonstructural prot
17	51	45.5	1196	2 J01467	toxin, nontoxic co
18	51	45.5	1196	2 S46430	botulinum neurotox
19	49	43.8	276	2 T33493	hypothetical prote
20	48	42.9	504	2 T47446	botulinum neurotox
21	48	42.9	1285	2 S70582	botulinum neurotox
22	48	42.9	1291	2 S46431	botulinum neurotox
23	48	42.9	1291	2 A49777	botulinum neurotox
24	47.5	42.4	1276	2 S11455	botulinum neurotox
25	47	42.0	359	2 F87937	protein F1486.6 li
26	47	42.0	385	2 T20879	hypothetical prote
27	47	42.0	489	2 B37837	probable alpha-amy
28	46	41.1	322	2 T25966	hypothetical prote
29	46	41.1	442	2 I47074	gene CD5 protein -

RESULT 1

BTCLTN

tentoxilysin (EC 3.4.24.68) precursor - Clostridium tetani

N;Alternate names: tetanus neurotoxin

C;Species: Clostridium tetani

C;Date: 31-Mar-1988 #sequence revision 31-Mar-1988 #text change 03-Jun-2002

C;Accession: A25689; A25757; A25194; B25194; B25194; A60759; S69348; S09364

R;Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, M.,

EMBO J. 5, 2495-2502, 1986

A;Title: Tetanus toxin: primary structure, expression in E. coli, and homology with botu

A;Reference number: A25689; MUID:87053814; PMID:3536478

A;Accession: A25689

A;Molecule type: DNA

A;Residues: 1-1315 <EIS>

A;Cross-references: GB:X04436; NID:940769; PIDN:CAA28033.1; PID:940770

R;Fairweather, N.F.; Lyness, V.A.

Nucleic Acids Res. 14, 7809-7812, 1986

A;Title: The complete nucleotide sequence of tetanus toxin.

A;Reference number: A25757; MUID:87040747; PMID:3774547

A;Accession: A25757

A;Molecule type: DNA

A;Residues: 1-1315 <FAI>

A;Cross-references: GB:X06214; NID:940773; PIDN:CAA29564.1; PID:940774

R;Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.

J. Bacteriol. 165, 21-27, 1986

A;Title: Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in E

A;Reference number: A25194; MUID:86085672; PMID:3510187

A;Accession: A25194

A;Molecule type: DNA

A;Residues: 743-1315 <FA2>

A;Cross-references: GB:M12739; NID:9144920; PIDN:AAA23282.1; PID:9144921

A;Accession: B25194

A;Molecule type: protein

A;Residues: 865-894 <FA3>

R;Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.

Infect. Immun. 57, 3588-3593, 1989

A;Title: Isolation, purification, and characterization of fragment B, the NH-2-terminal t

A;Reference number: A60759; MUID:90035436; PMID:2478476

A;Accession: A60759

A;Molecule type: protein

A;Residues: 461-475 <MAT>

R;Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.

J. Immunol. 142, 394-402, 1989

A;Title: Delineation of several DR-restricted tetanus toxin T cell epitopes.

A;Reference number: JS0098; MUID:89093918; PMID:2463305

A;Contents: annotation; epitope region

R;Schlavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R.

Nature 359, 832-835, 1992

A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyti

A;Reference number: S27125; MUID:93063293; PMID:1331807

A;Contents: annotation

R;de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.; Montecucco, C.
Eur. J. Biochem. 229, 61-69, 1995
A;Title: Structural studies on the zinc-endopeptidase light chain of tetanus neurotoxin.
A;Reference number: S69348; MUID:95262688; PMID:7744050
A;Accession: S69348
A;Molecule type: protein
A;Residues: 2-31 <DEF>
C;Comment: The source of this protein was an extrachromosomal plasmid.
C;Comment: The precursor is cleaved by endogenous proteinase activity to form light (fragment A) and heavy (fragment B) chains. The amino end of the heavy chain (fragment B) forms ion channels in a lipid bilayer. Fragment C binds to ganglionic presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of synaptobrevin.
A;Description: blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide bond in synaptobrevin.
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-457/Product: tentoxylisin light chain (fragment A) #status predicted <TTL>
F;461-1315/Product: tentoxylisin heavy chain (fragment B.C) #status experimental <TTH>
F;461-864/Domain: channel forming (fragment B) #status predicted <TXB>
F;865-1315/Domain: ganglioside binding (fragment C) #status predicted <TXC>
F;233,237/Binding site: zinc (His) #status predicted
F;234/Active site: Glu #status predicted

Query Match 100.0%; Score 112; DB 1; Length 1315;
Best Local Similarity 100.0%; Pred. No. 3.7e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVPKVSASHLE 21
Db 947 FNNFTVSFWLRVPKVSASHLE 967

RESULT 2
S33411
botulinum neurotoxin type F - Clostridium baratii
C;Species: Clostridium baratii
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999
C;Accession: S33411; S31860
R;Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T.
FEMS Microbiol. Lett. 108, 175-182, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium baratii type F neurotoxin.
A;Reference number: S33411; MUID:93252228; PMID:8486245
A;Accession: S33411
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1268 <THO>
A;Cross-references: EMBL:X68262; NID:g49138; PIDN:CAA48329.1; PID:g49139
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 55.4%; Score 62; DB 2; Length 1268;
Best Local Similarity 64.3%; Pred. No. 0.073;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVPK 14
Db 922 YQNFSISFWVRPK 935

RESULT 3
S48110
neurotoxin type F - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S48110
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinum neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48110
A;Status: preliminary; translation not shown
A;Molecule type: DNA

A;Residues: 1-366 <CAM>
A;Cross-references: EMBL:X70821; NID:g407792; PIDN:CAA50152.1; PID:g407793
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 54.5%; Score 61; DB 2; Length 366;
Best Local Similarity 57.1%; Pred. No. 0.028;
Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVPK 14
Db 297 YQNFSISFWVRPK 310

RESULT 4

S48109

neurotoxin type F - Clostridium botulinum (fragment)

C;Species: Clostridium botulinum

C;Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 16-Jul-1999

C;Accession: S48109

R;Campbell, K.D.; Collins, M.D.; East, A.K.

J. Clin. Microbiol. 31, 2255-2262, 1993

A;Title: Gene probes for identification of the botulinum neurotoxin gene and specific id

A;Reference number: S48103; MUID:94013372; PMID:8408542

A;Accession: S48109

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-369 <CAM>

A;Cross-references: EMBL:X70820; NID:g407790; PIDN:CAA50151.1; PID:g407791

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993

C;Superfamily: tetanus toxin

Query Match 54.5%; Score 61; DB 2; Length 369;
Best Local Similarity 57.1%; Pred. No. 0.029;
Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVPK 14
Db 297 YQNFSISFWVRPK 310

RESULT 5

I40813

neurotoxin type F - Clostridium botulinum

C;Species: Clostridium botulinum

C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Jul-1999

C;Accession: I40813; S48108

R;East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson, D.E.

FEMS Microbiol. Lett. 96, 225-230, 1992

A;Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.

A;Reference number: I40644

A;Accession: I40813

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1274 <RES>

A;Cross-references: GB:M92906; NID:g144866; PIDN:AAA23263.1; PID:g144867

R;Campbell, K.D.; Collins, M.D.; East, A.K.

J. Clin. Microbiol. 31, 2255-2262, 1993

A;Title: Gene probes for identification of the botulinum neurotoxin gene and specific id

A;Reference number: S48103; MUID:94013372; PMID:8408542

A;Accession: S48108

A;Status: preliminary; translation not shown

A;Molecule type: DNA

A;Residues: 634-1002 <CAM>

A;Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789

C;Superfamily: tetanus toxin

C;Keywords: neurotoxin

Query Match 54.5%; Score 61; DB 2; Length 1274;
Best Local Similarity 57.1%; Pred. No. 0.11;
Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVPK 14

Best Local Similarity 53.8%; Pred. No. 0.19; Mismatches 5; Indels 1; Gaps 0;

QY 1 FNNFTVSFWLRVP 13
Db 297 YKNFSISFWVRIP 309

RESULT 11
JH0256
botulinum neurotoxin type E precursor - Clostridium butyricum
C:Species: Clostridium butyricum
C>Date: 30-Jun-1992 #sequence_revision 15-May-1998 #text_change 16-Jul-1999
C:Accession: JH0256; S16145
R:Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A>Title: Sequences of the botulinum neurotoxin E derived from Clostridium botulinum type
A:Reference number: JH0256; MUID:92181428; PMID:1543481
A:Accession: JH0256
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-27,'E',29-1251 <POU>
A:Cross-references: EMBL:X62089; NID:940379
A:Experimental source: strains ATCC 43181 and ATCC 43755
R:Fujii, N.; Kimura, K.; Yashiki, T.; Indoh, T.; Murakami, T.; Tsuruki, K.; Yokosawa, N.
J. Gen. Microbiol. 137, 519-525, 1991
A>Title: Cloning of a DNA fragment encoding the 5'-terminus of the botulinum type E toxin
A:Reference number: S16145; MUID:91237316; PMID:2033376
A:Accession: S16145
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-229,'M',231-252 <FU>
A:Cross-references: EMBL:X53180; NID:940408
A:Experimental source: strain BL6340
C:Comment: The clostridial neurotoxins are toxins that inhibit neurotransmitter release
C:Superfamily: tetanus toxin
C:Keywords: neurotoxin
F;2-422/Product: botulinum neurotoxin type E light chain #status predicted <LIG>
F;423-1251/Product: botulinum neurotoxin type E heavy chain #status predicted <HEA>
F;412-426/Disulfide bonds: #status predicted

Query Match 50.0%; Score 56; DB 2; Length 1251;
Best Local Similarity 53.8%; Pred. No. 0.72;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVP 13
Db 912 YKNFSISFWVRIP 924

RESULT 12
S21178
botulinum neurotoxin type E precursor - Clostridium botulinum
C:Species: Clostridium botulinum
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Oct-1999
C:Accession: S21178; S48107; JH0257; B35294; A60027; S18111
R:Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.
Eur. J. Biochem. 204, 657-667, 1992
A>Title: The complete amino acid sequence of the Clostridium botulinum type-E neurotoxin
A:Reference number: S21178; MUID:92174922; PMID:1541280
A:Accession: S21178
A:Molecule type: DNA
A:Residues: 1-1252 <WHE>
A:Cross-references: EMBL:X62683; NID:940397; PIDN:CAA44558.1; PID:940398
R:Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A>Title: Gene probes for identification of the botulinum neurotoxin gene and specific id
A:Reference number: S48103; MUID:94013372; PMID:8408542
A:Accession: S48107
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 616-982 <CAM>

A:Cross-references: EMBL:X70815; NID:9407786; PIDN:CAA50146.1; PID:9407787
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
R:Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A>Title: Sequences of the botulinum neurotoxin E derived from Clostridium botulinum type
A:Reference number: JH0256; MUID:92181428; PMID:1543481
A:Accession: JH0257
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-176,'R',178-197,'C',199-339,'R',341-772,'I',774-962,'FE',965-966,'R',968-10
A:Cross-references: EMBL:X62089; NID:940393; PIDN:CAA43959.1; PID:940394
A:Experimental source: strain Beluga
R:Binz, T.; Kurazono, H.; Willie, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A>Title: The complete sequence of botulinum neurotoxin type A and comparison with other
A:Reference number: A35294; MUID:90264400; PMID:2160960
A:Accession: B35294
A>Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-176,'R',178-252 <BIN>
A:Experimental source: strain Beluga
R:Gimenez, J.A.; DasGupta, B.R.
Biochimie 72, 213-217, 1990
A>Title: Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the sit
A:Reference number: A60027; MUID:90344918; PMID:2116911
A:Accession: A60027
A:Molecule type: protein
A:Residues: 420-427 <GIM>
A:Experimental source: strain Beluga
A>Note: this fragment was generated by proteolysis with Lys-C rather than with trypsin
C:Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neu
C:Superfamily: tetanus toxin
C:Keywords: neurotoxin
F;1-422/Product: botulinum neurotoxin type E light chain #status predicted <LCH>
F;423-1251/Product: botulinum neurotoxin type E heavy chain #status predicted <HCH>
F;412-426/Disulfide bonds: #status predicted

Query Match 50.0%; Score 56; DB 2; Length 1252;
Best Local Similarity 53.8%; Pred. No. 0.72;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVP 13
Db 912 YKNFSISFWVRIP 924

RESULT 13
I40645
botulinum neurotoxin type A - Clostridium botulinum
C:Species: Clostridium botulinum
C>Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C:Accession: I40645
R:Willems, A.; East, A.K.; Lawson, P.A.; Collins, M.D.
Res. Microbiol. 144, 547-556, 1993
A>Title: Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A
A:Reference number: I40645; MUID:94143603; PMID:8310180
A:Accession: I40645
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1296 <RES>
A:Cross-references: EMBL:X73423; NID:9507070; PIDN:CAA51824.1; PID:9507071
C:Superfamily: tetanus toxin
C:Keywords: neurotoxin

Query Match 50.0%; Score 56; DB 2; Length 1296;
Best Local Similarity 50.0%; Pred. No. 0.75;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVPK 14
Db 938 YENFSTSFWRIPK 951

RESULT 14

S23158
 nucleocapsid protein - Impatiens necrotic spot virus
 C/Species: Impatiens necrotic spot virus
 C/Date: 22-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 20-Sep-1999
 C/Accession: S23158
 R/de Haan, P.; de Avila, A.C.; Kormelink, R.; Westerbroek, A.; Gielen, J.J.L.; Peters, P.
 FEBS Lett. 306, 27-32, 1992
 A/Title: The nucleotide sequence of the S RNA of Impatiens necrotic spot virus, a novel
 A/Reference number: S23158; MUID:92331780; PMID:1385787
 A/Accession: S23158
 A/Status: preliminary
 A/Molecule type: nucleic acid
 A/Residues: 1-449 <HAA>
 A/Cross-references: EMBL:X66972; NID:960488; PIDN:CAA47382.1; PID:960489
 A/Note: the sequence from Fig. 3 is inconsistent with that from Fig. 1 in having 104-Val
 C/Superfamily: tomato spotted wilt virus nonstructural protein NS
 C/Keywords: nucleocapsid

Query Match 46.4%; Score 52; DB 2; Length 449;
 Best Local Similarity 45.0%; Pred. No. 1.1;
 Matches 9; Conservative 5; Mismatches 4; Indels 2; Gaps 1;

QY 2 NN--FTVSFWLRVPKVSASH 19

DB 240 NNKPKISLWLRIPKIMKSN 259

RESULT 15

MNVUWC
 nonstructural protein NS - tomato spotted wilt virus (strain CPNH1)
 C/Species: tomato spotted wilt virus
 C/Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 16-Jun-2000
 C/Accession: JQ0547
 R/de Haan, P.; Wagemakers, L.; Peters, D.; Goldbach, R.
 J. Gen. Virol. 71, 1001-1007, 1990
 A/Title: The S RNA segment of tomato spotted wilt virus has an ambisense character.
 A/Reference number: JQ0547; MUID:90264829; PMID:1693160
 A/Accession: JQ0547
 A/Molecule type: genomic RNA
 A/Residues: 1-464 <DEH>
 A/Cross-references: DDBJ:D00645; NID:g222685; PIDN:BAA00540.1; PID:g222686
 C/Genetics:
 A/Gene: NS
 A/Map position: segment S
 C/Superfamily: tomato spotted wilt virus nonstructural protein NS
 C/Keywords: glycoprotein; nonstructural protein
 F;132,210,270,291,381/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 46.4%; Score 52; DB 1; Length 464;
 Best Local Similarity 64.3%; Pred. No. 1.2;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 NNFTVSFWLRVPKV 15

DB 239 SHFKLSLWLRVPKV 252

Search completed: July 22, 2003, 08:12:44
 Job time : 2.05732 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:05:44 ; Search time 91.3788 Seconds
(without alignments)
1830.070 Million cell updates/sec

Title: US-09-806-703A-4

Perfect score: 6812

Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTFAENPEYGLDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6812	100.0	1255	21	AAV92620 Human heregulin 2
2	6812	100.0	1255	22	AAE12130 Human tyrosine kin
3	6812	100.0	1255	22	AAE60167 HER2 transgene pla
4	6812	100.0	1255	23	AAU74545 Human HER2 (erbB2)
5	6806	99.9	1255	17	AAW01111 HER-2/neu protein.
6	6806	99.9	1255	20	AAW92406 Human HER-2/neu on
7	6806	99.9	1255	21	AAE21198 Human HER-2/neu pr
8	6806	99.9	1255	21	AAE84780 Amino acid sequenc
9	6806	99.9	1255	22	AAE85458 Human HER-2/neu pr
10	6806	99.9	1255	22	AAE88267 HER2/neu amino aci

11	6806	99.9	1355	23	AAE24067 Human Her-2 protei
12	6806	99.9	1355	23	AAE20479 Human Her-2/neu pr
13	6806	99.9	1355	23	AAE51143 Human Her-2/neu on
14	6806	99.9	1355	23	AAU77114 Human Her-2/neu po
15	6763	99.3	1433	14	AAE39568 Sequence of c-erbB
16	6642	97.5	1223	23	AAU98923 Human breast cance
17	6489	95.3	1300	21	AAE21208 Human HER-2/neu pr
18	5998.5	88.1	1356	21	AAE21199 Rat HER-2/neu prot
19	5998.5	88.1	1356	23	AAE51144 Mouse Her-2/neu on
20	5973.5	87.7	1356	21	AAE21206 Mouse Her-2/neu pr
21	5973.5	87.7	1356	22	AAE62860 Amino acid sequenc
22	5973.5	87.7	1356	23	AAE51151 Mouse Her-2/neu on
23	4892	71.8	919	21	AAE21203 Human HER-2/neu fu
24	4892	71.8	919	23	AAE51148 Her-2/neu extracel
25	4123.5	60.5	920	23	AAE51152 Mouse Her-2/neu ex
26	4123.5	60.5	926	23	AAE51153 Human HER-2/neu ex
27	3776	55.4	712	21	AAE21204 Human HER-2/neu fu
28	3776	55.4	712	23	AAE51149 Her-2/neu extracel
29	3630	53.3	782	18	AAW19764 Her2-GM-CSF immuno
30	3628	53.3	653	21	AAE21200 Extracellular HER-
31	3628	53.3	653	23	AAE51145 Human ErbB2 oncopr
32	3590	52.7	645	22	AAE60408 Human ErbB2 extrac
33	3590	52.7	645	22	AAE61593 DC8ecFv-erbB2EC fu
34	3525	51.7	951	21	AAE44993 Extracellular port
35	3422	50.2	624	11	AAE08222 Amino acid sequenc
36	3168	46.5	1210	21	AAE19259 Human EGF receptor
37	3168	46.5	1210	21	AAE50616 Human Her-2/neu on
38	3168	46.5	1210	23	AAE23019 Human Her-1 protei
39	3168	46.5	1210	23	AAE50768 Human epidermal gr
40	3166	46.5	1210	22	AAE68420 Amino acid sequenc
41	3127	45.9	1210	23	AAE51768 Human epidermal gr
42	3110.5	45.7	654	21	AAE21205 Rat HER-2/neu prot
43	3110.5	45.7	654	23	AAE51150 Rat Her-2/neu onco
44	3084	45.3	583	23	AAE20483 Human protein for
45	3084	45.3	587	23	AAE20481 Human protein for

ALIGNMENTS

RESULT 1
AAV92620
ID AAV92620 standard; Protein; 1255 AA.
XX
AAV92620;
AC
XX
DT 10-AUG-2000 (first entry)
XX
DE Human heregulin 2 (Her2).
XX
KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 1..173
FT /label= N-terminal
FT /note= "mature polypeptide"
FT Region 5..25
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Region 59..73
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Region 103..117
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Region 149..163
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Domain 174..323

Qy	* 781	YVSELLGICLTSTVQVLTQLMYPYGCILLDHYVRENKRGISQDLLNWCQIAKGMSYLSDVR	840
Db	781	YVSKLLGICLTSTVQVLTQLMYPYGCILLDHYVRENKRGISQDLLNWCQIAKGMSYLSDVR	840
Qy	841	LVHRDLAARNVLKVS PNHVKITD FGLARLLDIDETEYHADGGKVP I KWMALESILRRFT	900
Db	841	LVHRDLAARNVLKVS PNHVKITD FGLARLLDIDETEYHADGGKVP I KWMALESILRRFT	900
Qy	901	HQSDVWSYGVTVMELMTFGAKPYDGI PAREI PDLLEKGERLP OPPICTIDVYIMVMKCMW	960
Db	901	HQSDVWSYGVTVMELMTFGAKPYDGI PAREI PDLLEKGERLP OPPICTIDVYIMVMKCMW	960
Qy	961	IDSECRPRPRELVSEFSRMDRDPORFVVI QNEDLGAS PLDSTFYRSLLEDDDMGDVLDA	1020
Db	961	IDSECRPRPRELVSEFSRMDRDPORFVVI QNEDLGAS PLDSTFYRSLLEDDDMGDVLDA	1020
Qy	1021	EYLVPOQGFCDPAPAGAGVMVHRRHSSTRS GGGDLTLGLEPSEEEAPRSP LAPSEG	1080
Db	1021	EYLVPOQGFCDPAPAGAGVMVHRRHSSTRS GGGDLTLGLEPSEEEAPRSP LAPSEG	1080
Qy	1081	AGSDVFDGDLGMGAAKGLQSLPHDPSPLORYSEDPVPLPSETDGVVAPLTCSPQPEYV	1140
Db	1081	AGSDVFDGDLGMGAAKGLQSLPHDPSPLORYSEDPVPLPSETDGVVAPLTCSPQPEYV	1140
Qy	1141	NQPDVREPQPS PREGPLPAARPAAGATLERA KTLSPGKNGVVKDVFAPGGAVENPEYLT PQ	1200
Db	1141	NQPDVREPQPS PREGPLPAARPAAGATLERA KTLSPGKNGVVKDVFAPGGAVENPEYLT PQ	1200
Qy	1201	GGAAPOHPHPPAPSPAFDNLIYYMDQDPPERGAP PSTPKGTPTAENP YLGLDVVP	1255
Db	1201	GGAAPOHPHPPAPSPAFDNLIYYMDQDPPERGAP PSTPKGTPTAENP YLGLDVVP	1255

RESULT 2	
AAE12130	
ID	AAE12130 standard; Protein; 1255 AA.
XX	
XX	AAE12130;
XX	
DT	18-DEC-2001 (first entry)
XX	
DE	Human tyrosine kinase-type receptor, HER-2.
XX	
KW	Therapeutic compound; major histocompatibility complex; vaccine;
KW	antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
KW	adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
KW	antigen presenting cell; human; tyrosine kinase-type receptor.

XX	Key	Location/Qualifiers
FH	Region	774..782
FT		/note= "Antigenic epitope"
XX	WO200168677-A2.	
XX	20-SEP-2001.	
XX	16-MAR-2001; 2001WO-US40328.	
XX	16-MAR-2000; 2000US-0527487.	
XX	(GENZ) GENZYME CORP.	
XX	Nicolette CA;	
XX	WPI; 2001-616284/71.	
XX	N-PSDB; AAD19731.	
XX	Novel synthetic therapeutic compound for inducing immune response and	
PT	for use in adoptive immunotherapy, has enhanced binding to major	
PT	histocompatibility molecules and enhanced immunoregulatory properties	
PT	-	

XX		Claim 4; Page 63-67; 69pp; English.
XX		The invention relates to synthetic therapeutic compounds (antigenic peptides) with enhanced binding to major histocompatibility complex (MHC) molecules and enhanced immunoregulatory properties relative to their natural counterparts. Compounds of the invention are useful for inducing an immune response in a subject and for use in adoptive immunotherapy. They are useful as components of anti-cancer vaccines and to expand immune effector cells that are specific for cancers characterised by expression of the breast cancer antigen, HER-2. Polynucleotides that encode peptides of the invention are useful as hybridisation probes and as primers for the detection of genes of gene transcripts that are expressed in antigen presenting cells (APCs), to confirm transduction of polynucleotides into host cells. The present sequence is human tyrosine kinase-type receptor, HER-2. Compounds of the invention are designed based on the HER-2 antigenic peptide (774-782).
XX	SQ	Sequence 1255 AA;
		Query Match 100.0%; Score 6812; DB 22; Length 1255;
		Best Local Similarity 100.0%; Pred. No. 0;
		Matches 1255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	1	MELAAACRGLLALLPPGAASVCTGTDMLKRLPASBETHLDMLRHLYQGCVVGNL 60
Dd	1	MELAAACRGLLALLPPGAASVCTGTDMLKRLPASBETHLDMLRHLYQGCVVGNL 60
Qy	61	EITYLPFNASLSFLDIIQEYGVGYLLAHNQVRQLQRIRI VRGTOLFDNYALAVLDNG 120
Dd	61	EITYLPFNASLSFLDIIQEYGVGYLLAHNQVRQLQRIRI VRGTOLFDNYALAVLDNG 120
Qy	121	DPLNNTPTVTGASPGLRELQRSLEILIKGGVLIO RNPQLCYQDTILWKDI FHKNNOLA 180
Dd	121	DPLNNTPTVTGASPGLRELQRSLEILIKGGVLIO RNPQLCYQDTILWKDI FHKNNOLA 180
Qy	181	LTLIDTNRGRACHPCSPMKSGRCWGESSEDCOSLTRTCAGGCARCKGPLPDDCCHQC 240
Dd	181	LTLIDTNRGRACHPCSPMKSGRCWGESSEDCOSLTRTCAGGCARCKGPLPDDCCHQC 240
Qy	241	AAGCTGPKHSDCLACLHNFHGSI CELHLCPALTYNTDTFESMENPEGRTFFGASC VTACP 300
Dd	241	AAGCTGPKHSDCLACLHNFHGSI CELHLCPALTYNTDTFESMENPEGRTFFGASC VTACP 300
Qy	301	YNYLSTDVGSCTLVCPLHNQEVTAEDGTORCEKC KPCARVCYGLGM EHLREVR VTSAN 360
Dd	301	YNYLSTDVGSCTLVCPLHNQEVTAEDGTORCEKC KPCARVCYGLGM EHLREVR VTSAN 360
Qy	361	IQEFACKKIIFSLA FLPSFGDPASNATAPLOEQL QVPETLEEITGVLIYS AWPD SL P 420
Dd	361	IQEFACKKIIFSLA FLPSFGDPASNATAPLOEQL QVPETLEEITGVLIYS AWPD SL P 420
Qy	421	DLVSVFONLVQIRIGRI LHHNGAYSILTLOGGI SWLGRLS RELG SGLALI HNH T HL CF VH TV 480
Dd	421	DLVSVFONLVQIRIGRI LHHNGAYSILTLOGGI SWLGRLS RELG SGLALI HNH T HL CF VH TV 480
Qy	481	PWDQLFRPNHQALLHTANPEDECGVEGL A CH QL CARG HCWGPG PT QC VNCS OF LR QE OC 540
Dd	481	PWDQLFRPNHQALLHTANPEDECGVEGL A CH QL CARG HCWGPG PT QC VNCS OF LR QE OC 540
Qy	541	VBEICRVLQGI PR YVNVARH CLPCH PCQPONGSV TC FG PEADOCVA CAHY KD PPFC VAR C 600
Dd	541	VBEICRVLQGI PR YVNVARH CLPCH PCQPONGSV TC FG PEADOCVA CAHY KD PPFC VAR C 600
Qy	601	PSGVKP DL SYMP I WKF PD DE GA CP CPI N CT HS CV D LD DK GC PAEQRAS PL TS IV SA VG 660
Dd	601	PSGVKP DL SYMP I WKF PD DE GA CP CPI N CT HS CV D LD DK GC PAEQRAS PL TS IV SA VG 660
Qy	661	ILLVVVLGVVFG IL IKRR OOK IR KY TM RL LO B ET EL VE BL TP SG AMP NQA OW RI KE TEL 720
Dd	661	ILLVVVLGVVFG IL IKRR OOK IR KY TM RL LO B ET EL VE BL TP SG AMP NQA OW RI KE TEL 720

QY 721 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRENTSPKANKELDEAYYVAGVSP 780
Db 721 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRENTSPKANKELDEAYYVAGVSP 780
QY 781 YVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRNRRGLSGDGLNWCQIAKMSYLEVDV 840
Db 781 YVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRNRRGLSGDGLNWCQIAKMSYLEVDV 840
QY 841 LVHRDLAARNVVKSPNHVKITDGLARLLDIDETEVHADGGKVPKMALESILRRFT 900
Db 841 LVHRDLAARNVVKSPNHVKITDGLARLLDIDETEVHADGGKVPKMALESILRRFT 900
QY 901 HQSDVMSYGYVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYIMVWKWM 960
Db 901 HQSDVMSYGYVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYIMVWKWM 960
QY 961 IDSECRPRFRELVSERFMRARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDGMDLVA 1020
Db 961 IDSECRPRFRELVSERFMRARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDGMDLVA 1020
QY 1021 EYLVPOQGFCDPAPAGGMVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
Db 1021 EYLVPOQGFCDPAPAGGMVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAGLQSLPDLTHPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGLQSLPDLTHPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NQPDVPRQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTPO 1200
Db 1141 NQPDVPRQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPPPAPSPADNLYYWDQPPPERGAPSTFKGTPTAENPEYLGLDVPP 1255
Db 1201 GGAAPQHPPPAPSPADNLYYWDQPPPERGAPSTFKGTPTAENPEYLGLDVPP 1255

RESULT 3

AAB60167
ID AAB60167 standard; Protein; 1255 AA.
AC AAB60167;
XX 03-APR-2001 (first entry)
XX HER2 transgene plasmid construct encoded protein.
XX Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
KW antibody.
XX Homo sapiens.
OS Synthetic.
XX WO200100244-A2.
XX 04-JAN-2001.
XX 23-JUN-2000; 2000WO-US17229.
XX 25-JUN-1999; 99US-0141316.
PR 16-MAR-2000; 2000US-0189844.
XX (GETH) GENENTECH INC.
XX Erickson S, Schwall R;
XX WPI; 2001-061962/07.
DR N-PSDB; AAF24297.
XX Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
XX conjugating the antibody to a maytansinoid -

PS Example 3; Fig 4; 92pp; English.
XX The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX Sequence 1255 AA;
SQ Query Match 100.0%; Score 6812; DB 22; Length 1255;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPETHLDMRLHLYQCGVQVGNL 60
Db 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPETHLDMRLHLYQCGVQVGNL 60
QY 61 ELTYLPTNASLSFLODIQEVQVLIHNOVQVPLQRLRIVRGTOFEDNVALAVLDNG 120
Db 61 ELTYLPTNASLSFLODIQEVQVLIHNOVQVPLQRLRIVRGTOFEDNVALAVLDNG 120
QY 121 DPLNNTTPVTGASPGSLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKQNQLA 180
Db 121 DPLNNTTPVTGASPGSLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKQNQLA 180
QY 181 LTLIDTNRSRACHSPCKSGRCWGESSEDCQSLTRTVCCAGCARCKPLPTDCCHEQC 240
Db 181 LTLIDTNRSRACHSPCKSGRCWGESSEDCQSLTRTVCCAGCARCKPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTAC 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTAC 300
QY 301 YNYLSTDVGSCTLVCPHLNQEVTAEADGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEADGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
QY 361 IOEFAGCKIKI FGS LAF LPESFDGPASNTAPLOPEQLQVFETLEEITGYLISAWPDSL 420
Db 361 IOEFAGCKIKI FGS LAF LPESFDGPASNTAPLOPEQLQVFETLEEITGYLISAWPDSL 420
QY 421 DLSVFONLOVIRGRILHNGAYSILTLQGLIGISWGLRSLRELGSGLALIHNNHLCFVHTV 480
Db 421 DLSVFONLOVIRGRILHNGAYSILTLQGLIGISWGLRSLRELGSGLALIHNNHLCFVHTV 480
QY 481 PWDQLFRNPQHALLHTANRPEDECYGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
Db 481 PWDQLFRNPQHALLHTANRPEDECYGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
QY 541 VEECRVLOGLPREYVNAHCLCHPECPQNGSVTCFGEADQCACAHYKDPFPCVARC 600
Db 541 VEECRVLOGLPREYVNAHCLCHPECPQNGSVTCFGEADQCACAHYKDPFPCVARC 600
QY 601 PSGVRPDLISYMPIWKFPPDEEGACOPCINCTHSCVDLDDKGPAPORASPLTSIVSAVVG 660
Db 601 PSGVRPDLISYMPIWKFPPDEEGACOPCINCTHSCVDLDDKGPAPORASPLTSIVSAVVG 660
QY 661 ILLVVVLGVVFGILIKRROQKIRKVTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRROQKIRKVTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 721 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRENTSPKANKELDEAYYVAGVSP 780
Db 721 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRENTSPKANKELDEAYYVAGVSP 780
QY 781 YVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRNRRGLSGDGLNWCQIAKMSYLEVDV 840
Db 781 YVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRNRRGLSGDGLNWCQIAKMSYLEVDV 840
QY 841 LVHRDLAARNVVKSPNHVKITDGLARLLDIDETEVHADGGKVPKMALESILRRFT 900
Db 841 LVHRDLAARNVVKSPNHVKITDGLARLLDIDETEVHADGGKVPKMALESILRRFT 900

Db * 841 LVHRDLAARNVLYKSPNHVKITDFGLARLLIDIDETEHADGGKVPKMALESILRRPT 900
 Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYIMVKCWM 960
 Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYIMVKCWM 960
 Qy 961 IDSECRPRFRELVSERMDARQRFVVIQNEIDLGRASPLDSTFYRSLLEDMDGLDVA 1020
 Db 961 IDSECRPRFRELVSERMDARQRFVVIQNEIDLGRASPLDSTFYRSLLEDMDGLDVA 1020
 Qy 1021 EYVLVPOQGFCCPDPAAGAGVHHRHSSTSSGGGDLTLGLEPSEEEAPRPLAPSEG 1080
 Db 1021 EYVLVPOQGFCCPDPAAGAGVHHRHSSTSSGGGDLTLGLEPSEEEAPRPLAPSEG 1080
 Qy 1081 AGSDVFDGDLGMAAGKLSLTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 Db 1081 AGSDVFDGDLGMAAGKLSLTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 Qy 1141 NOPDVRPQPSREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTPQ 1200
 Db 1141 NOPDVRPQPSREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTPQ 1200
 Qy 1201 GGAAPQHPPPAFAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
 Db 1201 GGAAPQHPPPAFAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 4
 AAU74545
 ID AAU74545 standard; Protein; 1255 AA.
 XX AAU74545;
 XX
 XX 23-APR-2002 (first entry)
 XX
 DE Human HER2 (ErbB2) polypeptide.
 XX
 KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;
 KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
 KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
 KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
 KW glial disorder; astrocytal disorder; hypothalamic disorder;
 KW glandular disorder; macrophagal disorder; epithelial disorder;
 KW stromal disorder; blastocoealic disorder; inflammatory disorder;
 KW angiogenic disorder; immunological disorder.
 XX
 OS Homo sapiens.
 XX
 XX US2002001587-A1.
 XX
 XX 03-JAN-2002.
 XX
 XX 16-MAR-2001; 2001US-0811123.
 XX
 XX 16-MAR-2000; 2000US-189844P.
 XX
 XX 05-OCT-2000; 2000US-238327P.
 XX
 XX (ERIC/) ERICKSON S.
 XX (SCHW/) SCHWALL R.
 XX (SLIW/) SLIWOWSKI M.
 XX
 XX Erickson S, Schwall R, Sliwowski M;
 XX WPI; 2002-163686/21.
 XX N-PSDB; ABK14058.
 XX
 XX Treating tumour characterised by overexpression of epidermal growth
 XX factor receptor, ErbB or cancer in mammal, comprises administering
 XX anti-ErbB antibody-maytansinoid conjugate to the mammal
 XX
 XX Example 3; Fig 7; 93pp; English.
 XX
 XX The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor
 CC receptor (ErbB) and does not respond or responds poorly, to treatment
 CC with an anti-ErbB antibody, comprising administering to the mammal an
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
 CC prostate and bladder, preferably breast cancer. The breast cancer is a
 CC metastatic breast cancer or an aggressive form of metastatic breast
 CC cancer which overexpresses ErbB2. The method is also useful for treating
 CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
 CC epithelial, stromal, blastocoealic, inflammatory, angiogenic and
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)
 XX polypeptide of the invention.
 XX
 Qy Sequence 1255 AA;
 Db
 Qy Query Match 100.0%; Score 6812; DB 23; Length 1255;
 Db Best Local Similarity 100.0%; Pred. No. 0;
 Qy Matches 1255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db
 Qy 1 MELAALCRWGLLALLPPGAASSTQVCTGDMKRLRSPASPETHLDMLRHLYQGCVVQGNL 60
 Db 1 MELAALCRWGLLALLPPGAASSTQVCTGDMKRLRSPASPETHLDMLRHLYQGCVVQGNL 60
 Qy 61 ELYLPTNASLSFLQDIEVQGVVLIHNVQVPLQRLRIVRGTLFEDNYALAVLDNG 120
 Db 61 ELYLPTNASLSFLQDIEVQGVVLIHNVQVPLQRLRIVRGTLFEDNYALAVLDNG 120
 Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLTORNPOLCYQDTILWKDIFHKNOLA 180
 Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLTORNPOLCYQDTILWKDIFHKNOLA 180
 Qy 181 LTLIDTNRSRACHPCSPMCKGSRGWESSEDCQSLTRTVCCAGCARCKGPLPTDCCHEQC 240
 Db 181 LTLIDTNRSRACHPCSPMCKGSRGWESSEDCQSLTRTVCCAGCARCKGPLPTDCCHEQC 240
 Qy 241 AGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
 Db 241 AGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
 Qy 301 YNYLSDVGSCTLVCPHNVQVTAEDGTORCEKCKPCARVCYGLGMEHLREVRAVTSAN 360
 Db 301 YNYLSDVGSCTLVCPHNVQVTAEDGTORCEKCKPCARVCYGLGMEHLREVRAVTSAN 360
 Qy 361 IOEFAGCKKIFGSLAPLPSFDGDPASNTAPLQPEQLVFETILEETGYLISAWPDSLP 420
 Db 361 IOEFAGCKKIFGSLAPLPSFDGDPASNTAPLQPEQLVFETILEETGYLISAWPDSLP 420
 Qy 421 DLSVFONLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGSGLALIHNTHLCFVHTV 480
 Db 421 DLSVFONLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGSGLALIHNTHLCFVHTV 480
 Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPGTQCVCNCSQFLRGQC 540
 Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPGTQCVCNCSQFLRGQC 540
 Qy 541 VBECEVRLQGLPZYVYNARHCLPCHPECPQNGSVTCFGEADQCVCAHYKDPFPCVARC 600
 Db 541 VBECEVRLQGLPZYVYNARHCLPCHPECPQNGSVTCFGEADQCVCAHYKDPFPCVARC 600
 Qy 601 PSGVKPDLISYMPIWKFPDEEGACQPCPINCSTHSCVDLDDKGCPCAEQASPLTSIVSAVVG 660
 Db 601 PSGVKPDLISYMPIWKFPDEEGACQPCPINCSTHSCVDLDDKGCPCAEQASPLTSIVSAVVG 660
 Qy 661 ILLVWVLGVVFGILIKRROQKIRKVTMRLLQETELVEPLTPSGAMPNQAQRILKETEL 720
 Db 661 ILLVWVLGVVFGILIKRROQKIRKVTMRLLQETELVEPLTPSGAMPNQAQRILKETEL 720
 Qy 721 RKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYMAGVGSF 780
 Db 721 RKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYMAGVGSF 780
 Qy 781 YVSRLLGICLTSTVQLVTQJLMPYGCILLDHVRENRLGSLQDILLNMCWQIAKMSVLELYR 840

Db 781 YVSRLLGICLTSTVQLVTPYGLDHLVRENRLGSGDGLLNCWQIAKMSYLEDDR 840
Qy 841 LVHRDLAARNVLKSPNHVITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYIMWKCMW 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYIMWKCMW 960
Qy 961 IDSECRPRFRELVSFSEFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGMDLDA 1020
Db 961 IDSECRPRFRELVSFSEFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGMDLDA 1020
Qy 1021 EYLVPOQGFCDPPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
Db 1021 EYLVPOQGFCDPPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
Qy 1081 AGSDVFDGLGMAAGLQSLPTHDPSPLOKYSDDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGLGMAAGLQSLPTHDPSPLOKYSDDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Qy 1141 NQPDVROPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKVFAFGGAVENPEYLTPQ 1200
Db 1141 NQPDVROPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKVFAFGGAVENPEYLTPQ 1200
Qy 1201 GGAAPQHPHPPAPSPAFDNLVYWDQDPPRGAPPSTFKGTPTAENPEYLGIDVPV 1255
Db 1201 GGAAPQHPHPPAPSPAFDNLVYWDQDPPRGAPPSTFKGTPTAENPEYLGIDVPV 1255

RESULT 5

AAW01111 standard; Protein; 1255 AA.
XX AC AAW01111;
DT 01-JAN-1997 (first entry)
XX DE HER-2/neu protein.
KW HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW breast cancer; ovary cancer; colon cancer; lung cancer;
KW prostate cancer; immunisation; tumour; vaccine; vector.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Domain 676..1255
FT /label= Intracellular domain
FT /note= "Claimed domain, useful for immunisation"
XX PN W09630514-A1.
XX PD 03-OCT-1996.
XX PF 28-MAR-1996; 96WO-US01689.
XX PR 31-MAR-1995; 95US-0414417.
XX PA (UNIW) UNIV WASHINGTON.
XX PI Cheever MA, Disis ML;
XX DR WPI; 1996-455361/45.
XX DR N-PSDB; AAT40739.
XX PT DNA encoding HER-2-neu poly-peptide(s) - used for prevention or
XX PT treatment of malignancies with which the HER-2/neu oncogene is
XX PT associated
XX PS Claim 2; Page 56-61; 71pp; English.

XX CC Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.
XX SQ Sequence 1255 AA;
Query Match 99.9%; Score 6806; DB 17; Length 1255;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1253; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKRLPASPTHLDMLRHLYQGCVVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKRLPASPTHLDMLRHLYQGCVVQGNL 60
Qy 61 ELYLPTNASLSFLQDIQEVQVLIHNNQVQVPLQRLRIVRGTOIFEDNYALVDNG 120
Db 61 ELYLPTNASLSFLQDIQEVQVLIHNNQVQVPLQRLRIVRGTOIFEDNYALVDNG 120
Qy 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRPOLCQVDTILWKDIIFKKNOLA 180
Db 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRPOLCQVDTILWKDIIFKKNOLA 180
Qy 181 LTLIDNTRSRACHPCSPMCKGSRGCESEDQSLTRTVTCAGGACRCKGPLPTDCCHEQC 240
Db 181 LTLIDNTRSRACHPCSPMCKGSRGCESEDQSLTRTVTCAGGACRCKGPLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPRGRTTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPRGRTTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IQEFAGCKKIFGSLAFPLPSFDGDPASNTAPLQPEQLQVFETLEETGYLISAMPDLSL 420
Db 361 IQEFAGCKKIFGSLAFPLPSFDGDPASNTAPLQPEQLQVFETLEETGYLISAMPDLSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLLTQGLGSLGLSLRELGLALIHNNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLLTQGLGSLGLSLRELGLALIHNNTHLCFVHTV 480
Qy 481 PWDOLFPRNPHOALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQCVCNCSQFLRGQEC 540
Db 481 PWDOLFPRNPHOALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQCVCNCSQFLRGQEC 540
Qy 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
Db 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
Qy 601 PSGVKPDLSPYMPIWKEPDEEGACQPCPINCTHSCVDLDDKGCPEORASPLTSIVSAVNG 660
Db 601 PSGVKPDLSPYMPIWKEPDEEGACQPCPINCTHSCVDLDDKGCPEORASPLTSIVSAVNG 660
Qy 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNOAQRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNOAQRILKETEL 720
Qy 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGP 780
Qy 781 YVSRLLGICLTSTVQLVTPYGLDHLVRENRLGSGDGLLNCWQIAKMSYLEDDR 840
Db 781 YVSRLLGICLTSTVQLVTPYGLDHLVRENRLGSGDGLLNCWQIAKMSYLEDDR 840

Db 772 LLGLCGSSQLQVLYQLPLGSLLDHVKQHRETLGPQLLLNMGVQIAKGMYYLBEHSMVJR 831
QY 839 DLARNVVKSPNHNKITDFGLARLLDIDETEHADGKGKVPFKWMALESILRRFTHQSD 898
Db 832 DLALRNVLKSPSQVQVADFGVADLLPPDDKQLLHSEAKTFIKWMALESIFHCKYTHQSD 891
QY 899 VWSYGVTVWELMTFGAKPYDGIAPAREIPDLLEKGERLPQPPICITIDVYMWKCMWIDSE 958
Db 892 VWSYGVTVWELMTFGAEPYAGLRLAETIPDLLEKGERLAQPOICTIDVYMWKCMWIDEN 951
QY 959 CRPRFELVSEFRMARDPQRFVVIQNEIDGPASPLDSTFYRSLLEDDDDMGDLVDAEYL 1018
Db 952 IRTFRELANEFTRMARDPPYLVIKRAS-GPQTP--PAAPSPVITTKEL-----QAE 1003
QY 1019 VPQGGFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLRPSSE----- 1062
Db 1004 EPEL-----DLDLLEAESEGLATSIGSALSPLTGT 1034
QY 1063 -----EAPRSLAPSEG-----AGSDVPDGLGMAAGKLGSLPHTDPSLQRYSEDP 1110
Db 1035 LTRPRGSQSLSPSSGYMPMNQSSLGEACLDLSAVLGGREQFSRPISLH-PIPRGR----- 1088
QY 1111 TVPLPSETGVY---APL-----TC-----SPOPE-----YVNOPDVVRPQSPREGP 1150
Db 1089 --PASESEGHVTCSEAELOEKVSVCRSRSSRSPRGRSDVAVHQRHSLLTPTVTLSP 1146
QY 1151 -----LPAAPAGATLRAKTLSP-KGNQVY-----KDVFAFGGAVENPEYLT 1192
Db 1147 GLEEDGNGVMPDTHLRGASSREGTLSSVGLSSVLGTEEDED-----EYEVYMN 1198
QY 1193 PQGGAAPQPHPP 1204
Db 1199 RKRGSP-PRPP 1209
RESULT 11
TVFVLV
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
N;Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C;Species: avian leukosis virus, ALV
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C;Accession: B00643; A00643
R;Nilsson, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1995
A;Title: C-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: B00643
A;Molecule type: mRNA
A;Residues: 1-698 <NIL>
A;Cross-references: GB:M10066; GB:M13881; NID:G211749; PIDN:AAA48763.1; PID:G211750
A;Note: in Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gal
C;Comment: This protein is synthesized as a gag-env-erbB protein.
C;Genetics:
A;Gene: gag-env-erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p
F;1-6/Product: gag protein (fragment) #status predicted <GAG>
F;7-59/Product: env protein (fragment) #status predicted <ENV>
F;60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F;194-459/Domain: protein kinase homology <KIN>
F;202-210/Region: protein kinase ATP-binding motif
F;229/Active site: Lys #status predicted

Query Match 26.18; Score 1766.5; DB 1; Length 698;
Best Local Similarity 52.24; Pred. No. 2.3e-65;
Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18;

QY 572 GPADQCVACAHYKDPDFCVARCPGVKPDLSYMPIWKFPDEGACQPCINCTHSCVDL 631
Db 60 GP--DHCWKAHFIDGPHCVKACPAGVLGENDIL-VWKYADANAVCOLCHPNCRTGCKGP 116
QY 632 DDKGCAPAEQASPLTSTVSAVV-GILLVVLGVVFGILIKRQOKTKYTMRLLOSTEL 690

Db 117 GLEGCP---NGSKTFSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRTLRLLOEREL 172
QY 691 VEPLTPSGAMPNQAMRILKETELRKVKVGLSGAFGTVYKGIWIPDGENVKVIPVAIKVLJR 750
Db 173 VEPLTPSGEAPNOAHLRIILKETEFKKVKVGLSGAFGTVYKGLWIPEGEKVKIPVAIKELR 232
QY 751 ENTSPKANKEIIDEAAYVMAVGCSPPVSRLLGLCLTSTVOLVQLMPYGCGLLDHVRNRR 810
Db 233 EATSPKANKEIIDEAAYVMAVDNPHVCRLLGLCLTSTVOLITQLMPYGCGLLDYIREHKN 292
QY 811 LGSQDLNWCMIQAKMSYLEVRLVHRDLAARNVLKSPNHNKITDFGLARLLDIDETE 870
Db 293 IGSQYLLNWCQIAKGMVLEERLVRDLAARNVLKTPQHVKITDFGLAKLLGDAEXE 352
QY 871 YHADGKVP:KMALESILRRRFTHQSDVMSYGVTVWELMTFCAPYDGIAPAREIPDLLE 930
Db 353 YHAEKGKVP:KMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIAPASEISSVLE 412
QY 931 KGERLPQPPICITIDVYMWKCMWIDSECRPPRELVSFESRMARDPQRFVVIQ-NEDLG 989
Db 413 KGERLPQPPICITIDVYMWKCMWIDADSRPKFERLIAEFKWARDPPRYLVIQGDHMH 472
QY 990 PASPLDSTFYRSLLEDDDDMGDLVDAEYLVPOQGFCDPAPGAGGMVHRRSSSTRSG 1049
Db 473 LFPSTDSKFTYRIMEEDMEDIVDAEYLVPHQGF-----NSPST----- 513
QY 1050 GGDLTILGLRPSSEAPRSP-----APSEAGSDVFDGLGMAAGKLGSLPHTDPSLQ 1104
Db 514 -----SRTELLSSLSATSNNSATCID-----RNGQGHVPVRESFVQ 550
QY 1105 RYSEDDTVPDPSET--DGIVAPLTCSPQPEYVQNDVVRPQPPSPREGPLPAAPAGATLE 1162
Db 551 RYSSDPTGNFLBESIDDDGFL-----PAPEYVQ--LMPKPS----- 585
QY 1163 RAKTSLSPGKGVVGVKVF-----AFGAVENPEYLTPOGGAAPQHPHPPAF 1207
Db 586 ----TAMVQNYNNISLTAISKLPMSRYQNSHSTAVDNPEYL-----NTNOSPLA 633
QY 1208 SPAFONLYYWDQ-----DPPE-----RGAPPSTFKGTPTAENPEYLGIDVP 1248
Db 634 KTVFESSPYWIOGNHQNLDNPDVQDDFLPNETKENGLLKVPAAENPEYLVRAAP 689
RESULT 12
TVYUHV
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)
C;Species: avian erythroblastosis virus
C;Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C;Accession: A00644; A38022
R;Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A;Title: The erbB gene of avian erythroblastosis virus is a member of the src gene fami
A;Reference number: A00644; MUID:84026539; PMID:6313229
A;Accession: A00644
A;Molecule type: DNA
A;Residues: 1-604 <YAM>
A;Cross-references: GB:K01216; NID:G209676; PIDN:AAA42400.1; PID:G209678
R;Debuire, B.; Henry, C.; Benaissa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Martin,
Science 224, 1456-1459, 1984
A;Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type of
A;Reference number: A38022; MUID:84223957; PMID:6328658
A;Accession: A38022
A;Molecule type: DNA
A;Residues: 1-28,'W',30-139,'F',141-145,'V',147-152 <DEB>
A;Cross-references: GB:F02006
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific
F;130-395/Domain: protein kinase homology <KIN>
F;138-146/Region: protein kinase ATP-binding motif
F;165/Active site: Lys #status predicted


```
QY 691 VEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLR 750
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 114 VEPLTPSGEAPNQAHLAILKETEFKVKVLGFGAGFTVYKGLMIPGEKVTIPVAIKELR 173
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 751 ENTSPKANKELIDRAYVMAGVSGPYVSRLLGICLTSTVQLVQIMPYGCLLDHVRENRR 810
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 174 EATSPKANKELIDRAYVMASVDNPHVCRLLGICLTSTVQLITQIMPYGCLLDVIREHKDN 233
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 811 LGSQDLNWCQIAKAGMSYLEVDVLRHDLAARNVLKSPNHVKITDPFGLARLLDDETE 870
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 234 IGSQYLLNWCQIAKAGNYLEERHLVHRDLAARNVLKTPQDVKITDPFGLAKQLGADEKE 293
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 871 YHADGGKVPKIKWMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLE 930
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 294 YHAEGGKVPKIKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSVLE 353
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 931 KGERLPQPPICTIDVYIMVWKCMIDSECRPRFRELVSFSPRMARDPQRFVVIQ-NEDLG 989
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 354 KGERLPQPPICTIDVYIMVWKCMSDADSRPKFRELIAEFKWARDPPRYLVIQGDERMH 413
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 990 PASPLDSTFVRSLEDDMGDLVDAEYLYPQOGFCFDPAPGAGGMVHRRHSSSTRSG 1049
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 414 LPSFTDSKFRYTLMEBEDMEDIVDAEYLYVPHOGFF-----NSPST--- 454
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1050 GGDLTLGLEPSEEEAPRSP-----APSEGAGSDVFDGLMGAAKGLQSLPTHDPSPLO 1104
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 455 -----SRTPLLSLSATSNNATNCIDRNG-----H----- 481
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1105 RYSEDPVPLPSETDGVVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGAT-LER 1163
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 482 -----PVREDGFL-----PAPEYVYVQ--LMPKPESTAMVQNIYNYISLTAISK 523
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1164 AKTLSPGKNGVVKDVFAGGAVENPEYL 1191
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 524 LPMDSRYQN-----SHSTAVDNPEYL 544
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Search completed: July 22, 2003, 09:07:45
Job time : 30.7967 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 22.9062 Seconds
(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-465-479-12
Perfect score: 6804
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGPTAENPEYGLDVPV 1255
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution..

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6631	97.5	1255	A24571	protein-tyrosine k
2	5851.5	86.0	1254	I48161	p-185 precursor -
3	5844	85.9	1260	TVRTNU	protein-tyrosine k
4	3121	45.9	1210	GQHUE	epidermal growth f
5	3087	45.4	1210	TVCHLV	epidermal growth f
6	3074.5	45.2	1223	A53183	epidermal growth f
7	2943.5	43.3	1308	A47253	epidermal growth f
8	2658	39.1	1166	S06142	protein-tyrosine k
9	2384.5	35.0	1342	A36223	kinase-related tra
10	2300.5	33.8	1339	JC4387	epidermal growth f
11	1766.5	26.0	698	TFVPLV	protein-tyrosine k
12	1703	25.0	604	TVVUHV	protein-tyrosine k
13	1647	24.2	544	S35745	protein-tyrosine k
14	1640	24.1	545	S00727	kinase-related tra
15	1623	23.9	540	B44776	protein-tyrosine k
16	1621	23.8	540	TVFEVB	protein-tyrosine k
17	1620.5	23.8	1330	GQFVE	epidermal growth f
18	1479	21.7	644	A36325	epidermal growth f
19	1292	19.0	1323	E88257	protein let-23 (im
20	1292	19.0	1374	S70712	protein-tyrosine k
21	1196	17.6	1369	S70713	protein-tyrosine k
22	1169	17.2	1717	A45558	epidermal growth f
23	1108	16.3	527	A42032	epidermal growth f
24	948.5	13.9	843	A27131	epidermal growth f
25	806.5	11.9	346	S13807	protein-tyrosine k
26	754.5	11.1	311	S13808	protein-tyrosine k
27	724	10.6	1363	T43220	insulin-like growth
28	696	10.2	1382	INHUR	insulin receptor p
29	695	10.2	1383	A36080	insulin receptor p

ALIGNMENTS

RESULT 1

A24571

protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N:Alternate names: c-erb-B-2 protein precursor; Kinase-related transforming protein erbB

C:Species: Homo sapiens (man)

C>Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999

C:Accession: A24571; A25491; A44188; B44188; I59509; I57622

R:Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;

Nature 319, 230-234, 1986

A>Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth f

A:Reference number: A24571; MUID:86118663; PMID:3003577

A:Accession: A24571

A:Molecule type: mRNA

A:Residues: 1-1255 <YAM>

A:Cross-references: GB:X03363; NID:g31197; PIDN:CAA27060.1; PID:g31198

R:Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.

Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A>Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epiderm

A:Reference number: A25491; MUID:86016729; PMID:2995967

A:Accession: A25491

A:Molecule type: DNA

A:Residues: 737-1031 <SEM>

A:Cross-references: GB:M11767; NID:g182163; PIDN:AAA35808.1; PID:g553282

R:Coussens, L.; Yang-Peng, T.-L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg, P.

Science 230, 1132-1139, 1985

A>Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromos

A:Reference number: A44188; MUID:86070181; PMID:2999974

A:Accession: A44188

A:Molecule type: DNA

A:Residues: 740-910 <COUL>

A:Cross-references: GB:M12036; NID:g183988; PIDN:AAA35978.1; PID:g183989

A:Accession: B44188

A:Molecule type: mRNA

A:Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>

A:Cross-references: GB:M1730; NID:g183986

R:King, C.R.; Kraus, M.H.; Aaronson, S.A.

Science 229, 974-976, 1985

A>Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A:Reference number: I59509; MUID:85272597; PMID:2992089

A:Accession: I59509

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 832-909 <REX>

A:Cross-references: GB:L29395; NID:g459807; PIDN:AAA35809.1; PID:g459808

R:Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987

A>Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptional

A:Reference number: I57622; MUID:87286898; PMID:3039351

A:Accession: I57622

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-191 <TAL>

A;Cross-references: GB:M16792; NID:G183983; PIDN:AAA58637.1; PID:G553332
 C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
 C;Genetics:
 A;Gene: GDB:ERBB2; NGL; HER-2
 A;Cross-references: GDB:120613; OMIM:164870
 A;Map position: 17q21.1-17q21.1
 A;Introns: 25/1; 75/3; 147/1; 883/3
 A;Note: the list of introns is incomplete
 C;Function:
 A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
 C;Superfamily: epidermal growth factor receptor; protein kinase homology
 C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
 inase
 F;1-21/Domain: signal sequence #status predicted <SIG>
 F;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
 F;22-653/Domain: extracellular #status predicted <EXT>
 F;70-304/Domain: EGF receptor extracellular domain repeat <EE1>
 F;395-605/Domain: EGF receptor extracellular domain repeat <EE2>
 F;654-675/Domain: transmembrane #status predicted <TM>
 F;676-1255/Domain: intracellular #status predicted <INT>
 F;718-983/Domain: protein kinase homology <KIN>
 F;726-734/Region: protein kinase ATP-binding motif
 F;68,124,187,259,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
 F;753/Active site: Lys #status predicted
 F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 97.5%; Score 6631; DB 1; Length 1255;
 Best Local Similarity 97.6%; Pred. No. 1.3e-263;
 Matches 1225; Conservative 6; Mismatches 24; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLRASPETHLDMRLHYQCQVVGNNL 60
 DB 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLRASPETHLDMRLHYQCQVVGNNL 60
 QY 61 ELTYLPTNASLFLQIOBQVGVYLIHNNQVQLRIRVGRQLFEDNNYALAVLNG 120
 DB 61 ELTYLPTNASLFLQIOBQVGVYLIHNNQVQLRIRVGRQLFEDNNYALAVLNG 120
 QY 121 DPLNNTTPTVGASPGGLRELQRLSLEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
 DB 121 DPLNNTTPTVGASPGGLRELQRLSLEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
 QY 181 LTLIDNRSRACHPCSPMKSGRCWGESSEDCQSILTRTVACGACARCKGPLTDCCHQC 240
 DB 181 LTLIDNRSRACHPCSPMKSGRCWGESSEDCQSILTRTVACGACARCKGPLTDCCHQC 240
 QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
 DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
 QY 301 YNLYSTDVGSCTLVCPHNNQVTAEDGTORCEKSKPCARVCYGLGMQVIKANSKFIGIT 360
 DB 301 YNLYSTDVGSCTLVCPHNNQVTAEDGTORCEKSKPCARVCYGLGMQVIKANSKFIGIT 360
 QY 361 ELEFAGCKKIFGSLAFPLSPDGPASNTAPLQPEQLVFTEBITGYLISAMPDSLP 420
 DB 361 IOEFAGCKKIFGSLAFPLSPDGPASNTAPLQPEQLVFTEBITGYLISAMPDSLP 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRSLRELGSGLALIHNNHLCFVHTV 480
 DB 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRSLRELGSGLALIHNNHLCFVHTV 480
 QY 481 PHDQLFQVIKANSKFITELECYEGEGGLACHQLCARGHCWGPFGTQCNCQFIRGQEC 540
 DB 481 PHDQLFQVIKANSKFITELECYEGEGGLACHQLCARGHCWGPFGTQCNCQFIRGQEC 540
 QY 541 VEECRVLQGLPREYNARHCLPCHPECPONGSVTCFGEADQCVACAHYKDPPEFCVARC 600
 DB 541 VEECRVLQGLPREYNARHCLPCHPECPONGSVTCFGEADQCVACAHYKDPPEFCVARC 600
 QY 601 PSGVKPDLISYMPIWKPFPDEGACPCPINCTHSCVDLDDKGPAPQBRASPLTSIYSAVVG 660

DB 601 PSGVKPDLISYMPIWKPFPDEGACPCPINCTHSCVDLDDKGPAPQBRASPLTSIYSAVVG 660
 QY 661 ILLVVVLGVGGLIKRQOKIRKYTMRLLOETVELVEPLTPSGAMPNQAOQRIKTEL 720
 DB 661 ILLVVVLGVGGLIKRQOKIRKYTMRLLOETVELVEPLTPSGAMPNQAOQRIKTEL 720
 QY 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLENTSPKANKEIIDEAIVMAGVSP 780
 DB 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLENTSPKANKEIIDEAIVMAGVSP 780
 QY 781 YVSRLLGLCLSTVOLVTQMLPYGCLLDHVRNRLGSLQDLLNMCQIAKMSYLEVDR 840
 DB 781 YVSRLLGLCLSTVOLVTQMLPYGCLLDHVRNRLGSLQDLLNMCQIAKMSYLEVDR 840
 QY 841 LVHRDLAARNVLKSPNHNKVTDFGLARLLDIDETEHADGGKVPDKWMALESILRRPT 900
 DB 841 LVHRDLAARNVLKSPNHNKVTDFGLARLLDIDETEHADGGKVPDKWMALESILRRPT 900
 QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPPICITIDVYMINVKCM 960
 DB 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPPICITIDVYMINVKCM 960
 QY 961 IDSECRPRFRELVSFESFMSMARDPQRFVVIQNEGLGPASPLDSTFYRSLLDDMGDLVDA 1020
 DB 961 IDSECRPRFRELVSFESFMSMARDPQRFVVIQNEGLGPASPLDSTFYRSLLDDMGDLVDA 1020
 QY 1021 BEYLVPQGGFFCPDPAAGCAGGVHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
 DB 1021 BEYLVPQGGFFCPDPAAGCAGGVHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
 QY 1081 AGSDVDFDGLGWAAGKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEV 1140
 DB 1081 AGSDVDFDGLGWAAGKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEV 1140
 QY 1141 NQPDVRPOPSPRSGPLPAARPAAGATLERAKTLSPGKNGVVDVFAFGAVENPEYLTPO 1200
 DB 1141 NQPDVRPOPSPRSGPLPAARPAAGATLERAKTLSPGKNGVVDVFAFGAVENPEYLTPO 1200
 QY 1201 GGAAPOHPHPAPSPADNLYWDDPPERCAPSTFKGTPTAENPEYGLDVPV 1255
 DB 1201 GGAAPOHPHPAPSPADNLYWDDPPERCAPSTFKGTPTAENPEYGLDVPV 1255

RESULT 2
 148161
 p-185 precursor - golden hamster
 C;Species: Mesocricetus auratus (golden hamster)
 C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
 C;Accession: I48161
 R;Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa,
 Gene 140, 251-255, 1994
 A;Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
 A;Reference number: I48161; MUID:94193007; PMID:7908275
 A;Accession: I48161
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-1254 <RES>
 A;Cross-references: GB:D16295; NID:G493236; PIDN:BA003801.1; PID:G747595
 C;Genetics:
 A;Gene: neu
 C;Superfamily: epidermal growth factor receptor; protein kinase homology
 C;Keywords: ATP
 F;718-983/Domain: protein kinase homology <KIN>
 F;726-734/Region: protein kinase ATP-binding motif

Query Match 86.0%; Score 5851.5; DB 2; Length 1254;
 Best Local Similarity 86.0%; Pred. No. 8.6e-232;
 Matches 1079; Conservative 61; Mismatches 114; Indels 1; Gaps 1;

QY 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLRASPETHLDMRLHYQCQVVGNNL 60
 DB 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLRASPETHLDMRLHYQCQVVGNNL 60

QY 61 ELTYLPTNASTLFLQDIQEVQGVLLIAHNOVQVPLQRLRIYRGTLQPEDNYALAVLDNG 120
Db 61 ELTYLIPANATLSFLQDIQEVQGVLLIAHNOVQVPLQRLRIYRGTLQPEDNYALAVLDNR 120
QY 121 DPLNNTTPTVTGASPGRLQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNOLA 180
Db 121 DPLDNVTTATGTPGRLQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNOLA 180
QY 181 LTLIDITNRSRACHPCSPMCKSGRCWGESSEDQSLTRTVCAAGCARCKGPLPTDCCHQC 240
Db 181 PVDIDITNRSRACHPCSPMCKSGRCWGESSEDQSLTRTVCAAGCARCKGPLPTDCCHQC 240
QY 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESNPNPEGRTYTFGASCVTAC 300
Db 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESNPNPEGRTYTFGASCVTAC 300
QY 301 YNVLSTDVSGCTLVCPNHNQVTAEDGTQRCCKSKPCARVCYGLGMQVYKANSKFIGIT 360
Db 301 YNVLSTDVSGCTLVCPNHNQVTAEDGTQRCCKSKPCARVCYGLGMQVYKANSKFIGIT 360
QY 361 ELEFAGCKKIFGSLAPLPEFDPGDPASNTAPLOPEQLQVFETLEBITGYLIYSAMPDSL 420
Db 361 IQEAFAGCKKIFGSLAPLPEFDPGDPASNTAPLOPEQLQVFETLEBITGYLIYSAMPDSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLOGIGISWLGRLSRELCSGLALHNNHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLOGIGISWLGRLSRELCSGLALHNNHLCFVHTV 480
QY 481 PWDQLFQVIRKANSKFIGITELCEVCBGLACHOLCARGHCWGPGTQCNCVSOFLRGBC 540
Db 481 PWDQLFQVIRKANSKFIGITELCEVCBGLACHOLCARGHCWGPGTQCNCVSOFLRGBC 540
QY 541 VEECRVLQGLPREYV 600
Db 541 VEECRVLQGLPREYVY 600
QY 601 PSQVXPDLSYMPYVY 660
Db 601 PSQVXPDLSYMPYVY 660
QY 661 ILLVVLGVVFGILLKRRQKIRKYMRLRLQETELVEPLTPSGAMPNQAWRIKTEL 720
Db 661 ILLVVLGVVFGILLKRRQKIRKYMRLRLQETELVEPLTPSGAMPNQAWRIKTEL 720
QY 721 RKVKVLGSGAGFTVYV 780
Db 721 RKVKVLGSGAGFTVYV 780
QY 781 YVSRLLGLCLTSTVQLVTLQMPYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 840
Db 781 YVSRLLGLCLTSTVQLVTLQMPYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 840
QY 841 LVHRDLAARNVLKSPNPKVITDFGLARLLDIDETEHADGKVPDKWMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNPKVITDFGLARLLDIDETEHADGKVPDKWMALESILRRFT 900
QY 901 HQSDVWSYGVTVWELMTGAKPYDGIPIAREIPDLLEKGERLPQPPICITDVMYVWKCM 960
Db 901 HQSDVWSYGVTVWELMTGAKPYDGIPIAREIPDLLEKGERLPQPPICITDVMYVWKCM 960
QY 961 IDSECRPRFRELVSFBSRMARDPQRFVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 1020
Db 961 IDSECRPRFRELVSFBSRMARDPQRFVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 1020
QY 1021 EYLVPQGGFFCPDPAPGAGGVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 1080
Db 1021 EYLVPQGGFFCPDPAPGAGGVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 1080
QY 1081 AGSDVDFGDLGMAAGKGLQSLPHTDPLQRYSEDTPLPSETDGYVAPLTCSPOPEYV 1140
Db 1081 AGSDVDFGDLGMAAGKGLQSLPHTDPLQRYSEDTPLPSETDGYVAPLTCSPOPEYV 1140
QY 1141 NOPDVRPQPPSPREGFLPAARPAAGATLBRATKLSFGKNGVVKDVPFAGGAVENPEYLP 1200

RESULT 3
TVRTNU

Db 1141 NOPEVRPOPPIITPBGPLPVPAGATLBRPKTLSPGKNGVVKDVTFCGAVENPEYLP 1200
QY 1201 GGAAPOPPHPPAFSPAFDNLNYYWDDPPERGAPSTFKGTPTAENPEYLGLDVVP 1255
Db 1201 GGSASQPH-PFALCPAFDNLNYYWDDPSERGSPPNFTFEGTPTAENPEYLGLDVVP 1254
A:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
C:Accession: A24562; A61204
R:Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A:Title: The new oncogene encodes an epidermal growth factor receptor-related protein.
A:Reference number: A24562; MUID:86118662; PMID:3945311
A:Accession: A24562
A:Molecule type: mRNA
A:Residues: 1-1260 <BAR>
A:Cross-references: EMBL:X03362; NID:G56745; PIDN:CAA27059.1; PID:G56746
R:Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen, R.
Carcinogenesis 12, 1975-1978, 1991
A:Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no 2-thiazolylformamide or N-methyl-N-nitrosourea.
A:Reference number: A61204; MUID:92035293; PMID:1682063
A:Accession: A61204
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 637-663, 'V', 665-702 <MAS>
A:Note: authors translated the codon GCA for residue 25 as Val
C:Genetics:
A:Gene: neu
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>
F:658-680/Domain: transmembrane #status predicted <TMN>
F:723-988/Domain: protein kinase homology <KIN>
F:731-739/Region: protein kinase ATP-binding motif
F:771,191,263,535,576,634/Binding site: carboxylate (Asn) (covalent) #status predicted
F:691/Binding site: phosphate (Thr) (covalent) #status predicted
F:758/Active site: Lys #status predicted
F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted
Query Match 85.9%; Score 5844; DB 1; Length 1260;
Best Local Similarity 86.2%; Pred. No. 1.8e-231;
Matches 1083; Conservative 51; Mismatches 121; Indels 2; Gaps 2;
QY 1 MELAALCRWGLLLALLPPGAASCTQCTGDMKLRASPETHLDMRLHLYOGCQVQGNL 60
Db 4 MELAAWCRWGLLLALLPPGAGTQCTGDMKLRASPETHLDMRLHLYOGCQVQGNL 63
QY 61 ELTYLPTNASTLFLQDIQEVQGVLLIAHNOVQVPLQRLRIYRGTLQPEDNYALAVLDNG 120
Db 64 ELTYVVPANASLFLQDIQEVQGVLLIAHNOVQVPLQRLRIYRGTLQPEDNYALAVLDNR 123
QY 121 DPLNNTTPTVT-GASPGGLREQLRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNOL 179
Db 124 DPQDNVAASTTGTPTGELREQLRLSLTEILKGGVLIQORNPOLCYQDMVWKFVFRKNOL 183
QY 180 ALTLIDITNRSRACHPCSPMCKSGRCWGESSEDQSLTRTVCAAGCARCKGPLPTDCCHQC 239
Db 184 APVDIDITNRSRACHPCSPMCKSGRCWGESSEDQSLTRTVCAAGCARCKGPLPTDCCHQC 243
QY 240 CAAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESNPNPEGRTYTFGASCVTAC 299
Db 244 CAAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESNPNPEGRTYTFGASCVTTC 303
QY 300 PYNVLSTDVSGCTLVCPNHNQVTAEDGTQRCCKSKPCARVCYGLGMQVYKANSKFIGI 359
Db 304 PYNVLSTDVSGCTLVCPNHNQVTAEDGTQRCCKSKPCARVCYGLGMQVYKANSKFIGI 363

360 TELEFAGCKKIFGSLAFIPESFDGDPASNTAPLQEQVFTLBEITGYLISAWPDSL 419
364 NVQEPDCKKIFGSLAFIPESFDGDPSSGIAPLRPEQLQVFTLBEITGYLISAWPDSL 423
420 PDLVSFQNLQVIRGRILHNGAVSLTQGLIGISWGLRSLRSLGSLALHNNTHLCFVHT 479
424 RDLVSFQNLRIIRGRILHNGAVSLTQGLIGISWGLRSLRSLGSLALHNNTHLCFVHT 483
480 PWDOLFRQYIKANSKFIGITILE-CVGEGLACHQLCARHCHWGPCTQVNCNSQFLRQ 538
484 PWDQLFRNPHOALLHSGNRPEDELCVSSGLVCSLCAHGHCHWGPCTQVNCNSHFLRQ 543
539 ECVEECRVLOGLPREVYNARHCLCHPECPQNGSVTCFGRPADQCAAHYKDPFCVA 598
544 ECVEECRVWGLPREVYSDKRLCPCHPECPQNSSETCFGRPADQCAAHYKDSSCVA 603
599 RCPGVKPDLSYMPIWKFPDEGACQPCPINTHSCVDLDDKGCPCPAERASPLTSIVSAV 658
604 RCPGVKPDLSYMPIWKFPDEGICQPCPINTHSCVDLDDKGCPCPAERASPTFIATV 663
659 VGILLVVLGVVFGILLIKRQKIRKYTMRLLOETELVEPLTPSGAMPNQOMRLKET 718
664 EGVLLFLILVVVGLILIKRQKIRKYTMRLLOETELVEPLTPSGAMPNQOMRLKET 723
719 ELRKVKVLGSGAFGYVIGIWIPIPDGENVKIPVAIKVLRNTPSPKANKELDEAYVMGVG 778
724 ELRKVKVLGSGAFGYVIGIWIPIPDGENVKIPVAIKVLRNTPSPKANKELDEAYVMGVG 783
779 SPVSRLLGICLTSTVQLVTLMPYGLLDHVRNRLGSGQDLNWCQIAKMSYLE 838
784 SPVSRLLGICLTSTVQLVTLMPYGLLDHVRNRLGSGQDLNWCQIAKMSYLE 843
839 VELVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALSIILRR 898
844 VELVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALSIILRR 903
899 FTHQSDVNSYGVTVWELMTFGAKPDGIPAREIPDLLEKGERLPPOPICTIDVYIMWKC 958
904 FTHQSDVNSYGVTVWELMTFGAKPDGIPAREIPDLLEKGERLPPOPICTIDVYIMWKC 963
959 WMIDSECRPRFRELVSFSEFMRDQRFVVIQNEGLDGPASPLDSTFYRSLLEDMDGDLV 1018
964 WMIDSECRPRFRELVSFSEFMRDQRFVVIQNEGLDGPSPMDSTFYRSLLEDMDGDLV 1023
1019 DAEYLVLPQGGFCFDPAPACGMVHHRSSSTRSGGDLTLGLEPSEEARPRPLAPS 1078
1024 DAEYLVLPQGGFCFDPAPACGMVHHRSSSTRSGGDLTLGLEPSEEARPRPLAPS 1083
1079 EGAGSDVFDGDLGMAAGLQSLPHTDPSLQRYSEDPVPLPSETDGVVAPLTCSPQPE 1138
1084 EGAGSDVFDGDLGMAAGLQSLPHTDPSLQRYSEDPVPLPSETDGVVAPLTCSPQPE 1143
1139 YVNPQDVRPQPPSPREGPLPAARAGATLERAKTILSPKNGVYKDVFAFGGAVENPEYLT 1198
1144 YVNPQDVRPQPPSPREGPLPAARAGATLERAKTILSPKNGVYKDVFAFGGAVENPEYLT 1203
1199 PGGGAAPQHPAPSPAFDNLVYWDQDPERCAPSTFKGTPTAENPEYLGLDVVP 1255
1204 PREGTASPSPAFSPAFDNLVYWDQDSESGQPPSPNPEGTPTAENPEYLGLDVVP 1260

RESULT 4
GHUE
epidermal growth factor receptor precursor - human
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C:Species: Homo sapiens (man)
C:Date: 15-Nov-1984 #sequence revision 27-Nov-1985 #ext change 11-Jun-1999
C:Accession: A00641; A25772; S30024; A38672; A00642; A3615; A23062; A05281; A60143; A33
R.Ullrich, A.; Cousens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.; Y
rg, P.H.
Nature 309, 418-425, 1984
A:Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression of

A:Reference number: A00641; MUID:84219729; PMID:6328312
A:Accession: A00641
A:Molecule type: mRNA
A:Residues: 1-1210 <ULL>
A:Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA25240.1; PID:g575924
A:Note: the authors translated the codon AAG for residue 540 as Aen
R.Ishii, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
A:Title: Characterization and sequence of the promoter region of the human epidermal grc
A:Reference number: A25772; MUID:85270438; PMID:2991899
A:Accession: A25772
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-29 <ISH>
A:Cross-references: GB:M11234; NID:g181981; PIDN:AA52370.1; PID:g553272
R.Haley, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ullrich, A.; Waterfield, M.
Oncogene Res. 1, 375-396, 1987
A:Title: The human EGF receptor gene: structure of the 110 kb locus and identification c
A:Reference number: S30024; MUID:88217333; PMID:3329716
A:Accession: S30024
A:Molecule type: DNA
A:Residues: 1-29 <HA2>
A:Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
R.Haley, J.D.; Waterfield, M.D.
J. Biol. Chem. 266, 1746-1753, 1991
A:Title: Contributory effects of de Novo transcription and premature transcript terminat
A:Reference number: A38672; MUID:91107677; PMID:1988448
A:Accession: A38672
A:Molecule type: DNA
A:Residues: 1-29 <HAL>
A:Cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271
A:Experimental source: carcinoma cell line A431-7
R.Xu, Y.; Ishii, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.; Mer
Nature 309, 808-810, 1984
A:Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RNAs
A:Reference number: A00642; MUID:84245835; PMID:6330563
A:Accession: A00642
A:Molecule type: mRNA
A:Residues: 'RCAMRA', 150-187, 'KSVIOAV', 195, 'M', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-321
, 798-799, 'TD', 802-811, 'R', 813-942 <XUY>
A:Experimental source: A431 human carcinoma cells, which have large numbers of EGF recep
R.Lin, C.R.; Chen, W.S.; Krueger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verma, I.
Science 224, 843-848, 1984
A:Title: Expression cloning of human EGF receptor complementary DNA: gene amplification
A:Reference number: A43615; MUID:84196372; PMID:6326261
A:Accession: A43615
A:Molecule type: mRNA
A:Residues: 713-964 <LIN>
A:Experimental source: epidermoid carcinoma cell line A431
R.Simmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.
Biochem. Biophys. Res. Commun. 124, 125-132, 1984
A:Reference number: A23062; MUID:85046483; PMID:6093780
A:Accession: A23062
A:Molecule type: mRNA
A:Residues: 1028-1210 <SIM>
R.Weber, W.; Gill, G.N.; Speiss, J.
Science 224, 294-297, 1984
A:Reference number: A05281; MUID:84172183; PMID:6324343
A:Accession: A05281
A:Molecule type: protein
A:Residues: 25-30, 'S', 32-51, 454-467 <WEB>
R.Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.
J. Biol. Chem. 260, 5205-5208, 1985
A:Title: Identification of residues in the nucleotide binding site of the epidermal growth
A:Reference number: A60143; MUID:85182650; PMID:2985580
A:Accession: A60143
A:Molecule type: protein
A:Residues: 740-744, 'X', 746-747 <RUS>
R.Mroczkowski, B.; Mosig, G.; Cohen, S.
Nature 309, 270-273, 1984
A:Title: ATP-stimulated interaction between epidermal growth factor receptor and superco
A:Reference number: A38023; MUID:84191554; PMID:6325948
A:Contents: annotation; receptor activity

A>Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
 R:Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.
 Cell 59, 33-43, 1989

A:Title: Functional independence of the epidermal growth factor receptor from a domain

A:Reference number: A33331; MUID:90003333; PMID:2790960

A:Contents: annotation; internalization signal

C:Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor

C:Genetics:

A:Gene: GDB:EGFR

A:Cross-references: GDB:120610; OMIM:131550

A:Map position: 7p12.3-7p12.1

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-1210/Product: EGF receptor #status predicted <MAT>

F:25-645/Domain: extracellular #status predicted <EXT>

F:390-600/Domain: EGF receptor extracellular domain repeat <EE1>

F:646-668/Domain: transmembrane #status predicted <TM>

F:669-1210/Domain: intracellular #status predicted <INT>

F:710-975/Domain: protein kinase homology <KIN>

F:718-726/Region: protein kinase ATP-binding motif

F:999-1046/Region: coated-pit mediated internalization signal.

F:1047-1210/Region: inhibitory

F:128,175,352,413,444,528,603/Binding site: carbohydrate (Asn) (covalent) #status predic

F:745/Active site: Lys #status experimental

Query Match 45.9%; Score 3121; DB 1; Length 1210;

Best Local Similarity 49.7%; Pred. No. 2.2e-120; Indels 108; Gaps 23;

Matches 629; Conservative 174; Mismatches 355;

11 LLLALLPPGAA--STQVCTGTDMLRLPASPTTHLDMLRLHYQGVQVQGNLELYLPTN 68

14 LLAALCPASRALEKKVCGTQSNKLTQLGTFFDHFLSLQRMFNCEVVLGNLEITYVQRN 73

69 ASLSFLQDIQVGVVLIHNOVRQVPLRIIVRGTLQFEDNYALAVLDNGDPLNNTTP 128

74 YDLSEFLKTIQVAGVYLIATNVERIPLNQLIIRGNMYENSYALAVLSNYD----- 126

129 VTGASPGGLREQLRLSLTEILKGGVLIQNPOLCYODTILKWDIIFHKNQLALTLIDTNR 188

127 ---ANKTGELPMRLNQLIHLGAVRFSNNPALCNVESIQWRDIYSSDFLSNNMDFQNH 183

189 SRACHPCSPMKSGRCWGSSESDCSLRTVCAAGCA-RCKGLPLTDCCHQCAGAGCTGP 247

184 LSGCQKCDSPNGSCWAGAEENCQKLTIIQAQCSGRCKGSPSDCCNQCAGCTGP 243

248 KHSDDLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACPYNYLSTD 307

244 RESDCLVCRKFRDEATCKDTPPLMNYPTTYQMDVNPBGKYSFGATCVKCPRYNVVTD 303

308 VGSCTLVCLPHNQVTAEDGTORCEKSKPCARVCYGLQMWYIKANSKFIGITELE-FAG 366

304 HGSCVRACGADSYEM-EEGVRKRCCKEGPCRKVCNGIGIGBFK-DSLSINATNIIKHFN 361

367 CKKIFGSLAFSPESFDGDPASNTAPLQEQLOVFLEITGLYISAMPDLSLPVQ 426

362 CTSISGLHLVAFRGDSFTHTPPLDQELDIKTKVKEITGLLIQWPNRTDLHAF 421

427 NLQVIRGRILHNGAVSLTIQGLISWGLRSLRELSGLALIHNNHLCFVHTVPWDQLF 486

422 NLEIIRGRKQHQFSLAVVLSNITSLGRLSLKEISDGVIIISGNKVLKYANTINWKKLF 481

487 ROYIKANSKFIGITELEVCVGEGLACHOLCARGHCGPGTCVNCQSFIRGOCVECRV 546

482 GTSGQTKIISNRGENSKATQGVCHALCSPPGCVGPEPRDCVSCRNVSRGECVDCKL 541

547 LOGLPREYVARHCLPCHPECOFGNSVTCFGEADQCACAHYKOPPCVACRCPGKVP 606

542 LEGEPREFVENSEICQHECPLQAMNITCTGRGPDNCIQAHYIDGPHCVKTCFAGVWG 601

607 DLSYMPPIWKPDEGACQPCPINCTHSCVDLDDKGCAPBQBRASPLTISVAVVG---ILL 663

602 ENNTL-VWKYADAGHVCHLCHPNCTYCTGPGLEGCPNTGPKIP--SIATGMVGALLLLL 658

664 VVVLGVVFGILIKRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAOIRILKETELRV 723

659 VVALGIG---LFMRRHIVKRTLRLLQERELVEPLTPSGEAPNQALLRILKETEFKKI 715

724 KVLGSGAFGTYYKGIWIPDGNVPIPAVKVLRNTPSKANKEILDEAYVWAGVSPYVS 783

716 KVLGSGAFGTYYKGIWIPDGEKVKIPVAIKELREATSPKANKEILDEAYVWASVDNPHVC 775

784 RLIGICLTSTVOLVTPOLMPYCLLDHVRNRRGLSGODLLANCMQIAKGMVSLDRLVH 843

776 RLIGICLTSTVOLITQMLPPCCLLDYVREHKDNGISQVLLNWCVOIAKGMVLEDRLLVH 835

844 RDLAARNVLKSPNHVKITDFGLARLDIDETEHADGGKVPPIKWMALLESILRRFTQHS 903

836 RDLAARNVLKTPQHKITDFGLAKLLGAEKEYHAEGKVPPIKWMALLESILHRIYTHQS 895

904 DVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICTIDVYIMVVKCWMIDS 963

896 DVMSYGVTVWELMTFGSKPYDGPASEISSILEKGERLPQPPICTIDVYIMVVKCWMIDA 955

964 ECRPRFRELVSFMRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDMDGLVDAEE 1022

956 DSRPKRELIIIEFSKMRMDPORYLVIQDERMHLPSPTDSNFYRALMDEEDMDVDVDAE 1015

1023 YLVPQQGFFCPDPAPAGAGMVHRRHRSSTSGGGDLTLGLEPSEEEAPRSPAPSEAG 1082

1016 YLIPQQGFF-----SSPSTSRTPLLSSLSAT 1041

1083 SDVFDGDLGMAAKGLQSLPHTDPSLPQRYSEDPTVPLPSET--DGYVAPLTCSQPEYV 1140

1042 SN--NSTVACIDRNGLOSCPIKEDSFQRYSSDPTGALTEDSIDDTFL-----PVPSYI 1093

1141 NOPVVRPQPPRSGPLPAAPAGATLERAKTSLGKNGVVKVDFAFGAVENPEYL-TP 1199

1094 NO-SVPRKPSAGVQNPVYHNLNP-----APSRDPHYQD--PHSTAVGNPVLNTV 1142

1200 OGGAAPQHPAPAPSPAFDNLVYMDQ-----DP-----PERGAPSTFKGTPTA 1243

1143 Q-----FTCVNSTFDSFAHWAQKSHQISLDNPDYQODFFPKKPKNGIFKGS-TA 1192

1244 ENPEYL 1249

1193 ENAEYL 1193

RESULT 5

A33183

epidermal growth factor receptor precursor - mouse

C:Species: Mus musculus (house mouse)

C>Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999

C:Accession: A33183; A43318; S24942; A28941; S45325; I49643

R:Luetteke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N.A.;

Genes Dev. 8, 399-413, 1994

A:Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor

A:Reference number: A53133; MUID:94170986; PMID:8125255

A:Accession: A53183

A:Molecule type: mRNA

A:Residues: 1-1210 <LUE>

A:Cross-references: GB:U93425

R:Avivi, A.; Lavi, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.

Oncogene 6, 673-676, 1991

A:Title: Comparison of EGF receptor sequences as a guide to study the ligand binding sit

A:Reference number: A43818; MUID:91232866; PMID:2030916

A:Accession: A43818

A:Molecule type: mRNA

A:Residues: 1-714 <AVI>

A:Cross-references: GB:X39698

R:Eisinger, D.P.; Serrero, G.

submitted to the EMBL Data Library, June 1992

A:Reference number: S24942

A:Accession: S24942

A:Molecule type: mRNA
A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>
A:Cross-references: EMBL:Z12608
R:Heigermann, G.J.; Gill, G.N.
J. Biol. Chem. 263, 13152-13158, 1988
A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated i
A:Reference number: A28941; MUID:88330814; PMID:3138233
A:Accession: A28941
A:Molecule type: protein
A:Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000; 1002-1009,
R:Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
submitted to the EMBL Data Library, April 1994
A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
A:Reference number: S45325
A:Accession: S45325
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-971, 'K', 973-1210 <VER>
A:Cross-references: EMBL:X78987; NID:g488830; PIDN:CAA55587.1; PID:g488831
R:Paria, B.C.; Das, S.K.; Andrews, G.K.; Dev, S.K.
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse b
A:Reference number: I49643; MUID:93126380; PMID:7678348
A:Accession: I49643
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 12-20, 22-132 <RES>
A:Cross-references: GB:I06864; NID:g193001; PIDN:AAAS3029.1; PID:g567201
C:Genetics:
A:Gene: EGFR
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop
F:1-24/Domain: signal sequence #status predicted <SIG>
F:648-670/Domain: transmembrane #status predicted <TM>
F:712-977/Domain: protein kinase homology <KIN>
F:680,695/Binding site: phosphate (Thr) (covalent) #status experimental
F:720-728/Region: protein kinase ATP-binding motif
F:697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental
Query Match 45.4%; Score 3087; DB 2; Length 1210;
Best Local Similarity 49.5%; Pred. No. 5-se-119;
Matches 630; Conservative 168; Mismatches 363; Indels 112; Gaps 25;
QY 11 LLLALLPPGAA--STOVCTGTDMKRLPASPEHLDMLRLHYOGCGOVGVGNLELYLPTN 68
DB 14 LLTALCAAGALAEKKVCQGTNRITQLGTFFDHFLSLQRMVNNCEVLGNLEITYVQRN 73
QY 69 ASISFLQDIOEVGYVLIHNOVROVPLQRLIRVGTQLFEDNYALAVLDNGDPLNNTTP 128
DB 74 YDLSFLKTIQEVAGYVLIATNVERIPLENLQIRNLYENTYALAILSN----- 124
QY 129 VTGASPGGLRELQRLSLTBILKGGVLIQRNPOLCYQDTILWKDI----FHKNNQLALTLI 184
DB 125 -YGNRTGLRELPMRNLIQELIGAVFNSNPILCNWDITQWRDIQVNFWSNMDL---- 180
QY 185 DYNRSRACHPCSPMKGSKCWGSESSDDCSLRTVTCAGGCA-RCKGPLPTDCHEOCAAG 243
DB 181 -QSHPSSCPCKDPCSPGNGSCWGGEENCKLTKIIICAQQCCHRCGRSPSDCHNCAAG 239
QY 244 CTGPKHSDGLACLFHNSHGI CELHCPALVYNTDTFESMPNPEGRVTFGASCVTACPNYK 303
DB 240 CTGPRESDLVLCQKFODEATCKDTCPPLMLNPTTYQMDVNPBGKYSFGATCKKCPRYN 299
QY 304 LSTDVGSCTLVCPHLNHQVTAEDGTQRCBKSKPCARVCYGLGMQVYKANSKFIGITELE 363
DB 300 VTDHGSVCYRACPDYIEV-EEDGIRKCKKCDGCKVCNGIGIFGFK-DTILSINATNFK 357
QY 364 -FAGCKKIFGSLAFLEPESFDGDPASNTAPLOPEQLQVFTLEITGYLIVISAWPDSLPDL 422
DB 358 HFKYCTAIGDHLILPVAFGKGSFTRTPPLDPRELEILTKVKEITGFLLIQAWPDNWTDL 417

QY 423 SVFQNLQVIRGRIHLNGAYSLTLOGLIGISWLGRLSRLBELSGSLALIHHTHLCFVHTVPW 482
DB 418 HAFENLEIIRGTRKQHQGQFSLAVGLNITSLGRLSKELISDGDVLIISGNRLCYANTINW 477
QY 483 DQLFRQYIKANSKFTGITTELECVGEGGLACHOLCARGHCWGPGPTQCVNCSQFLRQECVE 542
DB 478 KKLFGTPNQKTKIMNRAEKCKAVNHVCNPLCSSEGCGPEPRDCVSCQVNSRGRECV 537
QY 543 ECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFPEADQCVACAHYKDPFCVACRPS 602
DB 538 KNLILEGPREFVENSECICQHEPLCPQAMNITCTGRGPDNCGICAHYIDGHPCHVKTCFA 597
QY 603 GVKPDLSPYMTWKFPEDEGACQPCPINCTHSCVDLDDKGCPEAEQASPLTSTVSVAVGIL 662
DB 598 GIMGENNTL-VWKYADANNVCHLCHANTYGCAGPLOGCEVWPSGPKIPSTATGIVGGL 656
QY 663 LVVLGVVFGI-LIKRQOKIRKYTMRLLOBELTELVEPLTPSGAMPNQAQMRILKETEUR 721
DB 657 LPIVV-VALGIGLFWRRRHIVKRTRLRLLOBELTELVEPLTPSGEAPNQAHLRIKETEPR 715
QY 722 KVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKELDEAYVMAGVGSY 781
DB 716 KIKVLGSGAFGTVYKGLWIPGEKVKIPVAIKELREATSPKANKELDEAYVMASVDNPH 775
QY 782 VSRLLIGICLTSTVQLVTLQMPYGCLLDHVRENRLGSLQDILLNWCWQIAKMSYLEDVRL 841
DB 776 VCRLLIGICLTSTVQLVTLQMPYGCLLDYVREHKNIGSQYLLNWCWQIAKGNLYLEDRL 835
QY 842 VHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILRRRPTH 901
DB 836 VHRDLAARNVLKTPQHVKITDFGLAKLLGAEKEYHAEGGKVPKIKWMALESILHRIYTH 895
QY 902 QSDWMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICTIDVYIMVKCMI 961
DB 896 QSDWMSYGVTVWELMTFGSKPYDGI PASDISILEKGERLPPOPICTIDVYIMVKCMI 955
QY 962 DSECRPRELVSFESRMARQPORFWIQQ-NEDLGPASPLDSTFYRSLLEDDDDMDGLVDA 1020
DB 956 DADSPKPRELILEFSQWARDQRYLVYQGDRLMPLPSTDSNFYALMDEEDMEDVDA 1015
QY 1021 EBYLVPQGGFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLPESEEAAPRSLAPSG 1080
DB 1016 DEYLPQGGFF-----NSPST-----SRTPLLSLS 1041
QY 1081 AGSDVFDGDLGMAAGKGLQSLTPHDPSPLOYSDEPTVPLPSET--DGVVAPLTCSQPE 1138
DB 1042 ATSN-----NSTVACINRNGSCRKVEDAFLOYSDDPTGAVTEDNIDDAFL-----PVPE 1091
QY 1139 VYNQPDVROPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKOVFAFGGAVENPEYL- 1197
DB 1092 VYNO-SVPRPAGSVQNPVYHNPQHP-----APGRDLHYQN--PHSNVAGNPEYL 1140
QY 1198 TPQGAAPQPHPPPPAFSPAFNLVYWDQ-----DP-----PERGAPSTFKGTP 1241
DB 1141 TAAQ-----PTCLSSGFNSPALWIKQSHOMSLDNPDYQDDFFPKETKENGIFKG-P 1190
QY 1242 TAENPEYLGLDVP 1254
DB 1191 TAENAEYLRVAPP 1203

RESULT 6
TVCHLV

epidermal growth factor receptor precursor - chicken
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C:Species: Gallus gallus (chicken)
C:Date: 28-Feb-1985 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
C:Accession: A27720; A00643
R:Jiax, J.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennstr
Mol. Cell. Biol. 8, 1970-1978, 1988
A:Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mol
A:Reference number: A27720; MUID:88261272; PMID:3260329

A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
A:Cross-references: GB:M20386
R:Nilsen, T.W.; Martone, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: A00643
A:Molecule type: mRNA
A:Residues: 585-1223 <NIL>
A:Cross-references: GB:M10066
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
specific protein kinase
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F:31-554/Domain: extracellular #status predicted <EXT>
F:81-307/Domain: EGF receptor extracellular domain repeat <EE1>
F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F:655-677/Domain: transmembrane #status predicted <TM>
F:678-1223/Domain: intracellular #status predicted <INT>
F:719-984/Domain: intracellular #status predicted <KIN>
F:727-735/Region: protein kinase ATP-binding motif
F:136,202,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #s
F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:754/Active site: Lys #status predicted
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat
Query Match 45.2%; Score 3074.5; DB 1; Length 1223;
Best Local Similarity 48.5%; Pred. No. 1.8e-118;
Matches 629; Conservative 173; Mismatches 349; Indels 147; Gaps 27;
8 RGLLLALLPGAA-----STQVCTGDMKRLPASPETHDMLRHLYQGCVQGNLE 61
13 RGAALVLLLLGVALCSAVEEKVCQGTNNKLTQLGHVDEHFTSLQRMNCEVLSNLE 72
62 LYLPLTNASLSFLODIQEVGVYLAHNOVROVPLQRLIRVGTOLFEDNYALAVLDNGD 121
73 IYVHNRLDTEFKTQEVAGVFLALNMDVPIENQLIRGNVLNDNSFALAVLSNVH 132
122 PLNNTTPVTGASPGGLRELQLRLSLTEILKGVLIQRNPOLCYQDTILMKDIFHKNNQAL 181
133 -NNKTQ-----GLRELPMKRLSEILNGVYKISNNPKLQNDIVLWNDIIDTSRK-PL 182
182 TLID-TNRSRACHPCSPCKGRCRCWSESSEDQSLTRTVACGCA-RCKGPLPTDCCHQ 239
183 TVLDFASNLSCPKCHPNCTEDHCWAGAGQNCQTLLTKVCAQCSGRCRGKVPSCCHNQ 242
240 CAAGCTGPKHSDCLACHENHSGICELHCPALVTYNTDTFESMPNPEGYTFGASCVTAC 299
243 CAAGCTGPKHSDCLACHENHSGICELHCPALVTYNTDTFESMPNPEGYTFGASCVTAC 302
300 PYNLTSDVGSCTVCLPLHNEVTAEDGTQRCCKSKPCARVCYGLGMVYKANSKFTGI 359
303 PHNVVYTHGSCVRSNCTDTEY-BENGVRCKCKCDGLCKSVKNGIGIGELAGILS-INA 360
360 TELE-PAGCKKIFGSLAFIPESFDGDPASNTAPLEQLQVETLEEITGYLYISAWPDS 418
361 TNIDSFKNCTKINGDVSILPFAFLGDAFTKTLPLDKLDVPTVKETISGFLLIQAWPDN 420
419 LPDLSVFQNLQVIRGLHNGAYSILTQGLGTSWIGLSRLRELGSGLALIHNTLCFVH 478
421 AFDLYAFENLEIIRTKQKHQGYSLAVNWLKIQSLGLRSLKISDGDIAIMKNKLVAD 480
479 TVPWQDLQFQYIKANSKFTIGITECEVCGELACHOLCARGHCWGPGPTQVCNCSOFLRGO 538
481 TNWNSLFAQSQKTKLIQNRNKNNDCTADRHVCDPLCSVDGCGWGPCHFCRCRFSRQK 540
539 ECVECRVLQGLPREYVNAHCLPCHPECPQNG----SVTCFGPEADQCACAHYKDPFF 595

RESULT 7

A47253
epidermal growth factor receptor, HER4 - human
C:Species: Homo sapiens (man)
C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C:Accession: A47253
R:Plowman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; Ne
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A:Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal
A:Reference number: A47253; MUID:93189574; PMID:8383326
A:Accession: A47253
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1308 <PILO>
A:Cross-references: GB:L07868; NID:g337359; PIDN:AAB59446.1; PID:g337360
A:Note: sequence extracted from NCBI backbone (NCBIP:126842)
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor
F:716-981/Domain: protein kinase homology <KIN>
F:724-732/Region: protein kinase ATP-binding motif

541 ECVKCNITLQGEPRFEDSKLCPCHSECLVQNSTAYNTTCSGPGDCKMCAHFIDGPH 600
596 CVARPCSGVKPDLSTYMPWKPDDEBGACQPCPINCTHSCVDLDDKGCPCAEQASPLTSTV 655
601 CVKACPAVGENDTL-VWKYADANAVCOLCHPNCTRCCKGPGLEGCP---NGSTPTSTA 656
656 SAVV-GILLVVVLGVVFGILIKRROOKIRKYTMRLLOETELVEPLTPSGAMPNOAQMRI 714
657 AGVGGLLCLLVVGLIGLGLYLR--HIVKRTLRLLQERELVEPLTPSGEAPNQAHLRI 715
715 LKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYM 774
716 LKETEFKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYM 775
775 AGVGSPPYVSRLLGICLTSTVQLITQLMYPYGLLDHVRNRRGLSGQDLLNWCQIAKGN 834
776 ASVDNPHVCRLLGICLTSTVQLITQLMYPYGLLDHVRNRRGLSGQDLLNWCQIAKGN 835
835 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESI 894
836 YLEERLVRDLAARNVLKTPQHVKITDFGLAKLIGADEKEYHAEGGKVPKIMMALESI 895
895 LRRPETHQSDVWSYGVTVWELMTFGSKPYDGPASEISSVLEKGERLPQPPICTIDVMI 954
896 LHRITYTHQSDVWSYGVTVWELMTFGSKPYDGPASEISSVLEKGERLPQPPICTIDVMI 955
955 MVKCMWIDSECRPRELVSFSSWARDPQRFVVIQ-NEDLGASPLDSTFVRSLEDDDD 1013
956 MVKCMWIDADSRPFRELIAEFKWARDPPRYLVIQGERMHLPSPTOSKFTVRLMBEED 1015
1014 MGLVDABEYLVPQQGFCCPDPAFGAGGMVHRRSSSTRSGGGDLTLGLPESEEAPRS 1073
1016 MEDIVDADEYLVPHQGF-----NSPST-----SRT 1041
1074 PL-----APSGAGSDVFDGLGMAAGKLOSLPHTDPSLPORVSEDTVPPLPST--DG 1126
1042 PLLSLSLATSNSATNCID-----RNGQGHFVREDSFVQYSSDPTGNFLEESIDDG 1093
1127 YVAPLTCSPQPEYVQNPQDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKGVVKDYF- 1185
1094 FL-----PAPEYVQ--LMPKFS-----TANQNIYINISL 1124
1186 -----AFGAVENPEYLTQGGAAQPQHPPPAFSPAFDNLVYDQ----- 1225
1125 TAISKLPMSYQNSHSAVDNPEYL-----NTNOSPLAKTVFESSPYIQSGNHQI 1176
1226 --DPPE-----RGAPPSTFKGTPTAENPEYLGLDVP 1254
1177 NLDNPDYQODFLPNETKENGLLKVPAENPEYLRVAAP 1214

Db 123 YQK-NPSSP--DYVQVGLKQLQLSNLTSLSGGVKVSHPNPLNCLNVETINWWDIVDKTNP 179
QY 180 ALTLIDTNRSRACHPCSPMKGSRGWGESSEDCQSLTRTVACAGC-ARCKGPLPTDCCHE 238
Db 180 TNNLIPHAPEQCQKCDHSCVNGSCWAPGPGHCQKFTKLLCAEQCNRRCRGPKPDCCNE 239
QY 239 QCAAGCTGPKHSDCLACLHFNHSGTICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTA 298
Db 240 HCAGCTGPRATDCLACRDFNDGCTKPTCPPKIYDIVSHQVVDNPNKIYTFGAACVKE 299
QY 299 CFYNYLSTDVSGCTLVCPLHNOEVTAEQTCRCEKSCPCARVCVGLGMOYIKANSKFTG 358
Db 300 CFSNVVTE-GACVRSNCAGMLEVD-ENGKRSCKPCDGVCPKVDGIGLSI-SMTIAVN 356
QY 359 ITEL-EFAGCKKIFGSLAPLPESFDGDPASNTAPLQEQLOVFETLEETIGLYISAMPD 417
Db 357 STNIRSFNCTKINGDIILNRNSFEGDPHYKIGTMDPEHLNLTITVKEITGYLVIMWPE 416
QY 418 SLPDLVSFQNLQVIRGRILHNGAYS-LTLQGLGISWGLRSURELGSGLALIHHTHLCF 476
Db 417 NMTLSVFNQLEIIRGRTTFRSGFVQVVRHLQWGLRSLSKEYVSAGNVILKNTLQRLY 476
QY 477 VHTVPWDOLFROYIKANSKFIGITELCEVGEGLACHOLCARGHCWPGPTOCVNCQFRL 536
Db 477 ANTIWRLFRSEDSQ-----IYDARTENQTCNNECSEDCW-PGTMCVSLHVD 528
QY 537 GQECVECKVLQGLPREYNARHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPC 596
Db 529 GGRCVASCNLLGEPREAVQDGRVCQHQECLVQDLSLTCYGPANCKSAHFQDGPQC 588
QY 597 VARCPSGVKPDLSYMPIKFPDEBAGACQPCPINCETHSCVDLDDKGPAPQASPLTISVS 656
Db 589 IPRCPHGILGDGDTL-INKYADKMGQCPCHQNCQCGSGPGLSGCRGD-IVSHSSLAGV 646
QY 657 AVUGILLVVVLGVVFGILIKRQKIRKYVWRLLOETVELPELTPSGAMPNQAQWRLK 716
Db 647 LVSGGLTIVALLIIVLURRRIR-KRKTIRCLLOEKLVELBPLTPSGQAPNQAFLRLK 705
QY 717 ETELKRVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEIILDEAYVMAG 776
Db 706 ETEFKKDRVLGSGAFGVYKGLWNPGENIRIPVAIKVLRNTPSKANKEIILDEAYVMAG 765
QY 777 VQSPVSRLLGLTSTVOLVTLMPYGLLDHVRNRRGLSGDQLLWNCQIAKGMVYL 836
Db 766 VDPHVCRLGLTCLTSAVOLVTLMPYGLLDYVVRQHERICQGLLWNCVQIAKGMVYL 825
QY 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLAELLDDITETVHADGCKVPIKWMALLESILR 896
Db 826 EERHLVHRDLAARNVLLKNPNHVKITDFGLSKLLTADEKEYQADGCKVPIKWMALLESILQ 885
QY 897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYMTMV 956
Db 886 WTYTHQSDVWSYGVTVWELMTFGSKPYDGIPIAREIPDLLEKGERLPQPPICTIEVYMLIL 945
QY 957 KCMWIDSECRPRFRELVSFESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDMDGD 1016
Db 946 KCMWIDPSRPRFRELVSFESQARDPSRYLVIQ--NLPSSLDRLLFSRLSSDD--D 1000
QY 1017 LVDAEYLVPOQFPCPDPAAGAGGVHRRSSSTRSGGDLTLGLPSEEEAPRSLA 1076
Db 1001 VVDAEYLLPYKRI-----NRQGS-----EPC 1023
QY 1077 PSEAGSDVFDGDLGWAAGLQSLPHTDPSPLQRYSEDPV-PLPSETDGYVAFLTCSP 1135
Db 1024 PPTGH-----PVRENSITLRSIDPTQNALEKLDGH----- 1055
QY 1136 QBEYVNPQDVRPQP-----PSPRE-----GPLP-AARPAGATLAKTILSPGKNGVVK 1182
Db 1056 --EYVNPQGETSSRLSDIYNPNYEDLTDGWPVSLSSQEAETNFSREYLTNTQNSL-- 1111
QY 1183 DVFAEGGAVENPEYLTPOGGAAPQHPPPAPSPAFDNLYYWDQDPERGAPPSTFKGPT 1242
Db 1112 -PLVSSGSMDDPDY---QAG-----YQAAF-----LPQTGALTNGMFLPLA 1148

QY 1243 AENPEYLG 1250
||| |||
Db 1149 AENLEYLG 1156

RESULT 9

A36223
kinase-related transforming protein (erbB3) (EC 2.7.1.-) precursor - human
C:Species: Homo sapiens (man)
C>Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C:Accession: A36223; I59164
R:Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A:Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal growth factor receptor gene family.
A:Reference number: A36223; MUID:90083234; PMID:2687875
A:Accession: A36223
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1342 <KRA>
A:Cross-references: GB:M29366
R:Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, G.J.
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A:Title: Molecular cloning and expression of another epidermal growth factor receptor-related gene, erbB-3.
A:Reference number: I59164; MUID:90311312; PMID:2164210
A:Accession: I59164
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-559, 'G', 561-957, 'P', 959-1063, 'G', 1065-1342 <RES>
A:Cross-references: GB:M34309; NID:G183990; PIDN:AAA35979.1; PID:G306841
C:Genetics:
A:Gene: GDB:ERBB3; HER3
A:Cross-references: GDB:119880; OMIM:190151
A:Map position: 12q13-12q13
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
C:Keywords: ATP; phosphotransferase
F:707-972/Domain: protein kinase homology <KIN>
F:715-723/Region: protein kinase ATP-binding motif

Query Match 35.0%; Score 2384.5; DB 2; Length 1342;
Best Local Similarity 40.0%; Pred. No. 2.8e-90;
Matches 529; Conservative 194; Mismatches 446; Indels 155; Gaps 35;

QY 10 GLLALLPPGAA--STQVCTGTMKRLPASPETHLMLRHLVQGCQVVOGNLELTLYPT 67
||| |||
Db 11 GLLFSLARGSEVGNQVAVCPGTLNGLSVTGDAENQYQTLKLYERCEVVMGNLEIVLGH 70
QY 68 NASLSFLQDIQEVQGYVLIHNOVRQVPLQRLIRVGTQLFEDNYVALVLDNGDPLNNTT 127
||| |||
Db 71 NADLSFLQWIEVTVGYVLYVAMNBFSTLPLNLRVVRGTQVYDGKFAIFVM-----LNNT 125
QY 128 PVTGASPGGLSELOLRSLTEILKGVLIQRPOLCYODTILWKDIFHKNNQLAULTIDTN 187
||| |||
Db 126 ----NSSHALKQLKTLQTEILSGGVIEKNKDKLCHMDTIDWRDLVRDRD----AEIVVKD 178
QY 188 RSRACHPCSPMKGSRGWGESSEDCQSLTRTVACAGC-ARCKGPLPTDCHEQCAAGCTG 246
||| |||
Db 179 NGRSCPPCHEYCKG-RWNGPGESEDCQTLTKTICAPQCNHCFGNPNQCHDEACAGGSG 237
QY 247 PKHSDCLACLHFNHSGTICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACPNYLYST 306
||| |||
Db 238 PQDTCFACRHFNDSGACVPRCPQPLVYVKNLTFLQLEPNPHTKYQYGVGVCAVSCPHNFV-V 296
QY 307 DVGSCTLVCPLHNOEVTAEQTCRCEKSCPCARVCVGLGMOYIKANSKF--IGITELE- 363
||| |||
Db 297 DQTSVCRACPPDKMEVD-KNGLKMCPECGGLCPKACEGTG-----SGSRFQTVDSNIDG 350
QY 364 FAGCKKIFGSLAPLPESFDGDPASNTAPLQEQLOVFETLEETIGLYISAWPDSLPLDS 423
||| |||
Db 351 FVNCTKILGNLDLITGLNGDPWHKIPALDPEKLNVTREITGYLNIQSWPHMNFPS 410
QY 424 VFQNLQVIRGRILHNGAYS-LTLQGLGISWGLRSRLRELGSGLALIHHTHLCFVHTVPW 482
||| |||

[illegible]

RESULT 10

RESOL
JC4387

UC438/
epidermal growth factor receptor homolog precursor - rat

N;Alternate names: ErbB3 protein; HER3 protein

C:Species: *Rattus norvegicus* (Norway rat)
N/Accline name: E1DB3 protein; HER3 p1

C/species: Rattus norvegicus (Norway rat)
C/Date: 17-Jan-1996 #sequence revision 19-Apr-1996 #text change 13-Nov-1998

C;Date: 17-Jul-1996
C;Accession: JC4387

C/ACCESSION: UC4387
R/Hellyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.

K, HELLYER, N.O.; KILM,
Gene 165, 279-284, 19

Gene 165, 279-284, 1995

A:Reference number: J54387; MUID:96096535; PMID:8522190
A:Accession: J54387
A:Molecule type: mRNA
A:Residues: 1-1339 <HEL>
A:Cross-references: GB:U29339; NID:G915389; PID:G915390
A:Experimental source: liver
A:Note: The authors translated the codon AAC for residue 369 as Thr and GTT for residue 370.
C:Comment: This protein is a functional heregulin receptor that transduces signals to th C;Genetics:
A:Gene: ErbB3
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog C;Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F:640-659/Domain: transmembrane #status predicted <TMW>
F:705-970/Domain: protein kinase homology <KIN>
F:713-721/Region: protein kinase ATP-binding motif
F:939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr) (C

Query Match	33.8%	Score 2300.5	DB 2	Length 1339
Best Local Similarity	40.3%	Pred. No. 7.5e-87		
Matches	520	Conservative 173	Mismatches 429	Indels 167
Gaps	37			
Qy	3	LAALCRGLLALLALPGAA---STQVCTGTDMKLRLPASPETHLDMLRHLYQGVQGN	59	
Db	7	LQVLC----FLLSLARGESENGNSQAVCPGFLNGLSVTGDADNQVQTYLKYLYKEVVMGN	62	
Qy	60	LELTYPNTASLFLQDIQISVOGVLIATHNQVQVPLQRLRIRVGTQLPEDNTALAVLDN	119	
Db	63	LEIVLTGHNADLSFLQWIREVATYAVLVANNEFSVLPPLNLRVVRGTQVYDGKFAIFVM--	120	
Qy	120	GDPINNTPVTGASPGGLRELQRLSLTEILKGVLQIRNPQLCYQDTILMKDIFHKNNQL	179	
Db	121	---LNYNT---NSSHALRQLKFTQLTEILSGGVYIEKNDKLCMDTIDWRDIVRVR--	170	
Qy	180	ALFLIDTNRSRACHPCSPMCKSGRCRWESSEDCQSLTRTVACGC--ARKGPIPTDCCHE	238	
Db	171	GAEIVKNNNGANGPPCHVECKG--RCWGGPGDDCQILTKTICAPQCNRCRFGPNQOCCHD	229	
Qy	239	QCAAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTYNTDTFESMPNPEGRYTFGASCVTA	298	
Db	230	ECAGCGSGPDOTDCAFRRNDSGAVPRCPPEPLVYTNKLTFLQELPHHTKIYGGVCVAS	289	
Qy	299	CPYNLSTDVGSCTLVCPHLENQEVTAEDGTQRCCKSKPCARVCYGL--GMQYIKANSKF	356	
Db	290	CPHNFV--VDQTFCVRACPPDKMEVD--KHGLKMCPECGGLCPKACEGTGSSRYQTVDSSN	347	
Qy	357	IGITELEFACCKIFGSLAPLPSFSGDPAASNTAPIQPEOLQVFTLEETLEETLYISAMP	416	
Db	348	ID----GFVNCTKILGNLDFLTGLNVDPWHKIPALDPEKLNVRFTVREITGYLNIQSWP	403	
Qy	417	DSLPLDSVFQNLQVIRGRILHNAGYS--LTLOGIGISWLGRLSRELGSGLALTHNTHLC	475	
Db	404	PHHNFSVFNLTITIGRSSLYNRGFSLLINKNLNVTSLGFRSLKEISAGRVIISANQQLC	463	
Qy	476	FVHTVPDQLFR--QYIKANSKE--IGITELECVGEGLACHOLCARHGCWGGPQTQCVN	530	
Db	464	YHSLNWTRLRPSERLDIKYDRPLG----ECLAEGKVCDFLCSGGCGWGPAGQCLS	519	
Qy	531	CSQFLRQGEVCEBRVLQGLPREYVYNAHCLPCHPECPQNGSVTFCGPBDAQCVACAHY	590	
Db	520	CRNYSREGVCVTHCNFLQGEPRFVHEAQFCSCHPECLPMEGISTYNGSSGDACARCAHF	579	
Qy	591	KDPPFCVARCPSGVKPDLSVMPYTKPFDEBEGACQPCINCTHSC--VDLDDKGPAPBQRA	648	
Db	580	RDGPHCVNSCPHGILG--AKGPIYKYPDAQNECRPCNCHENTQGCNGPELQDCLGQAEVLM	637	
Qy	649	SPLTSTIVSAVVGLVVVLGVWVGILIKRQKIR--KYTMRRLLQETELVEPLTPSSGAMP	707	
Db	638	SKPHLVIAVTVG--LAVILMILGGSLFYWRGRRIQNRAMRRYLRGESLEPLDPS--EKA	694	
Qy	708	NOAQMRILKETELRKVKVLGSGAFGTGYKGIWTPDGENVKIPVAIKVLRENTSPPKANKEI	767	

Db 695 NKVLARIFKETELRLKLVLSGVFGTVHKGIWIPGESIKIPVICIKVIEDKSGRSQFQAV 754
QY 768 LDEAYMAGVGPYSRLLGICLTSTVQLVTOIMPYGCILLDHVHNRGRGLSGQDLINWCM 827
Db 755 TDHMLAVGSLDHAHIVRLGLCPGSSQLVTVQLPLGSLLDHVQKHRETLPQLLNWGV 814
QY 828 QIAGKMSYLEDVLRHDLAARNVLKSPNHVKITDFGLARLLDDIDETEHADGKGVPIK 887
Db 815 QIAGKMYLLEHSMVHRDLAARNVLMKSPSQVADFGVADLLPDKQLLHSEAKTPIK 874
QY 888 WMALSIILRRRTHQSDVMSYGVTVWELMTFGAKPYDGLPAREIPDLLEKGERLPOPPIC 947
Db 875 WMALSIHFGKTHQSDVMSYGVTVWELMTFGAEPYAGLRALAEIPDLLEKGERLAQPOIC 934
QY 948 TIDVYMWKWMIDSECRPRFRELVSFMRMARDPQRFVVIQNEDELGPASPLDSTFYRS 1007
Db 935 TIDVYMWKWMIDENIRPTFKELANEFTRMARDPPRYLVIKRAS-GPGTP--PAAEPS 991
QY 1008 LLEDMDGLVDAAEYLVPOQGFPCDPAPGAGGVHHRHRSSTRSGGDDLTILGLEPSE 1067
Db 992 VLTTEL-----QEALEPEL-----DLDLLEABE 1017
QY 1068 E-----EAPRSPLAPSEG-----AGSDVFDGLGMAAKGLQ 1099
Db 1018 EGLATSLGALSILPTGLTRPGSQSLSPSSGYMPMNQSSLGEACLDASVILGREGQFSR 1077
QY 1100 SLPTHPSPFLORYSEDTVPLPSETDGYV-----APL-----TC-----SPOPE-----Y 1139
Db 1078 PISLH-PIPRGR-----PASESEGHVGTGSEAELEQKVSVCRSRSRSRSPRGDSAY 1129
QY 1140 VNQPDVRPOPSPREGP-----LPAARPAGATLERAKTILSP-GKNGV----- 1181
Db 1130 HSQRHSLTPVPLSPGPLEEDGNGYVMPDTHLRGASSREGTLLSVGLSSVLGTTEED 1189
QY 1182 KDVFAGGAVENPEVLTPOGGAAPOPHP 1210
Db 1190 ED-----EYEYINWRKRGSP-PRPP 1209
RESULT 11
TVFVLV
Protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
N:Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C:Species: avian leukosis virus, ALV
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C:Accession: B00643; A00643
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: B00643
A:Molecule type: mRNA
A:Residues: 1-698 <NID>
A:Cross-references: GB:M10066; GB:M13881; NID:g211749; PIDN:AAA48763.1; PID:g211750
A>Note: In Genbank entry CHKGRBBF, release 109.0, the source is designated as Gallus gal
C:Comment: This protein is synthesized as a gag-env-erbB protein.
C:Genetics:
A:Gene: erbB
A:Accession: A00644
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
F:1-6/Products: gag protein (fragment) #status predicted <GAG>
F:7-59/Product: env protein (fragment) #status predicted <ENV>
F:60-69/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F:194-459/Domain: protein kinase homology <KIN>
F:202-210/Region: protein kinase ATP-binding motif
F:229/Active site: Lys #status predicted
Query Match 26.0%; Score 1766.5; DB 1; Length 698;
Best Local Similarity 52.2%; Pred. No. 2.5e-65;
Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18;
QY 578 GPEADQCVAHYKDPFPFCVACRCPGVKPDLSYMPKPPDEGACQPCPINCTHSCVDL 637

Db 60 GP--DHCWKCAHFDGPHCVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTGCKGP 116
QY 638 DDKGPABQBPASPLTSIVSAVV-GILLVVLGVVFGILIKRRQKIRKYIMRRLQETEL 696
Db 117 GLEGCP--NGSKTPSIAAGVVGGLCLVVGVLGILYLRRL-HIVRKRTLRLRLQEREL 172
QY 697 VEPLTPSGANPQAMRILKETELRKVKVLSGAPGTYYKGIWIPDGENVKIPVAKVL 756
Db 173 VEPLTPSGEAPQAHRLILKETEFKVKVLSGAFGTYYKGLWIPEGEKVKIPVAKEL 232
QY 757 ENTSPKANKEILDEAYVMAGVSPVSRLLGLICLTSTVQLVTOIMPYGCILLDHVHNRGR 816
Db 233 EATSPKANKEILDEAYVMASVDNPHVCRLLGLICLTSTVQLITQLIMPYGCILLDYIIRHKON 292
QY 817 LGSQDLNWCQIAKMSYLEDVLRHDLAARNVLKSPNHVKITDFGLARLLDIDETE 876
Db 293 IGSQVLLNWCQIAKGMNLEERRLVHRDLAARNVLVKTPOHVKITDFGLAKLILGADAKE 352
QY 877 YHADGGKVPILKWMALLESILRRFPTHQSDVMSYGVTVWELMTFGAKPYDGLPAREIPDLLE 936
Db 353 YHAEKGKVPILKWMALLESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGLIPASEISSVLE 412
QY 937 KGERLPOPPICITIDVYMWKWMIDSECRPRFRELVSFMRMARDPQRFVVIQ-NEDLG 995
Db 413 KGERLPOPPICITIDVYMWKWMIDADSRRFKRELIAEFKRNARDPPRYLVIOQDERMH 472
QY 996 PASPLDSTFYRSLLDDEDDMDGLVDAAEYLVPOQGFPCDPAPGAGGVHHRHRSSTRSG 1055
Db 473 LPSPDTSKFYRLMEEDMEDIVDAEYLVPHQGF-----SRTPLLSSLSATSNNSATNCID-----NSFST--- 513
QY 1056 GGDLTGLLEPSEBAPRPL-----APSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPQL 1110
Db 514 -----SRTPLLSLSATSNNSATNCID-----RNGQGHVPREDVSFVQ 550
QY 1111 RYSEDPVPLPSET--DGYVAPLTCSPQPEYVQNDPVRPQPSREGPLPAARPACATLE 1168
Db 551 RYSSDPTGNFLEESIDDGFL-----PAPEYVNO--LMPKKPS----- 585
QY 1169 RAKTILSPGKNGVVKDVF-----AFGAVENPEVLTPOGGAAPQHPHPPAF 1213
Db 586 -----TAMVQNIYNNISLTAISKLPMSRYQNSHSTAVDNPEYL-----NTNQSLA 633
QY 1214 SPAFDNLVYWDQ-----DPPE-----RGAPPSTFKGTPTAENPEYILGLDVP 1254
Db 634 KTVFESSPYWTQSGNHQINLDNPYQDPLFNKPNGLLKVPAAENPEYILRVAAP 689
RESULT 12
TVYUHH
Protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)
C:Species: avian erythroblastosis virus
C:Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C:Accession: A00644; A38022
R:Iyamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A:Title: The erbB gene of avian erythroblastosis virus is a member of the src gene fami
A:Reference number: A00644; MUID:84026539; PMID:6313229
A:Accession: A00644
A:Molecule type: DNA
A:Residues: 1-604 <YAM>
A:Cross-references: GB:K01216; NID:g209676; PIDN:AAA42400.1; PID:g209678
R:Debuire, B.; Henry, C.; Benatissa, M.; Bisette, G.; Claverie, J.M.; Saule, S.; Martin,
Science 224, 1456-1459, 1984
A:Title: Sequencing the erba gene of avian erythroblastosis virus reveals a new type of
A:Reference number: A38022; MUID:84223957; PMID:6328658
A:Accession: A38022
A:Molecule type: DNA
A:Residues: 1-28,'W',30-139,'F',141-145,'V',147-152 <DEB>
A:Cross-references: GB:K02006
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogens; phosphotransferase; transforming protein; tyrosine-specific p

F:130-395/Domain: protein kinase homology <KIN>
F:138-146/Region: protein kinase ATP-binding motif
F:165/Active site: Lys #status predicted

```
Query Match      25.0%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 8.5e-63;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

QY 587 CAHYKDPPEVCARCPGKPDLSYMPIMKFPDEEGACQPCPNCNTHSCVDLDDKCPAQ 646
Db 3 CAHFDGPHCVKACFAGVLGENDTL-VKRYADANAVCQCHNCNTRGCKGPGLEGCP--- 58

QY 647 RASPLTSIVSAVV-GILLVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGA 705
Db 59 NSGKTPSIAAGVVGGLLVVGLGIGLYLRR-HIVKRTLRLLQERELVEPLTPSGE 117

QY 706 MNQAMRILKETELRKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLRNENTSPKANK 765
Db 118 APNQAHLRIKETEFKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLRNENTSPKANK 177

QY 766 EILDAYMAGVSPVSRLLGICLTSTVQLTQLMYPYGLLDHVRNRRGLSGSOLLNW 825
Db 178 EILDAYMNASVDNPHVCRLLGICLTSTVQLTQLMYPYGLLDYIREHKNDIGSYLLNW 237

QY 826 CMQIAKMSYLEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEYHADGKVP 885
Db 238 CVQIAKMWYLERLVRDLAARNVVKTPQHVKITDFGLAKLGADEKEYHAEGKVP 297

QY 886 IKWMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQP 945
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSVLEKGERLPQP 357

QY 946 ICTIDVYIMVKWMDSECRPRFELVSEFMRMDPQRFVVIQ-NEDLGPASLDSTF 1004
Db 358 ICTIDVYIMVKWMDIDASRKPRFELIAEFKSMARDPPRYLVIQGDERRMLPSTDSKF 417

QY 1005 YRSLLDEDDMGDLVDAEYLVPPQGFCCPDAPGAGGMVHRHRSSTRSGGDLTLGLE 1064
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449

QY 1065 PSEERAPRSL-----APSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPQRYSDPTVP 1119
Db 450 -----SRTPLLSLSATSNNATNCID-----RNQGHFVREDSPVQRYSSDPTCN 495

QY 1120 LPSET--DGVAPLTCSQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLPCK 1177
Db 496 FLEESIDGFL-----PAPEYVNO--LMPKKPSTAM----- 524

QY 1178 NGVWKDVAF-----GGAVENPEYLTQCGAAPPHPPPAFSPAFD 1218
Db 525 --VQNIYNFISLTALSKLPMSRYQNSHSTAVDNPAYL-----NTNQSPKATVPE 574

QY 1219 NYIYWDQDPPERGAPSTFKGTPTAENPEY 1248
Db 575 SSPYIYQSGNHQ-----INLDNPDY 594
```

RESULT 13

S35745

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
C:Species: avian erythroblastosis virus

C>Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997

C:Accession: S35745

R.Vennstrom, B.

submitted to the EMBL Data Library, March 1993

A:Reference number: S35743

A:Accession: S35745

A:Molecule type: DNA

A:Residues: 41-544 <VEN>

A:Cross-references: EMBL:X12707

C:Genetics:

A:Gene: erbB

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific
F:135-400/Domain: protein kinase homology <KIN>
F:143-151/Region: protein kinase ATP-binding motif
F:170/Active site: Lys #status predicted

```
Query Match      24.2%; Score 1647; DB 2; Length 544;
Best Local Similarity 54.9%; Pred. No. 1.5e-60;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

QY 578 GPEADQCVACAHYKDPPEVCARCPGKPDLSYMPIMKFPDEEGACQPCPNCNTHSCVDL 637
Db 1 GP--DHCMKCAHFDGPHCVKACFAGVLGENDTL-VWKYADANAVCQCHNCNTRGCKGP 57

QY 638 DDKGPAPORASPLTSIVSAVV-GILLVVLGVVFGILIKRQOKIRKYTWRRLLQETEL 696
Db 58 GLEGCP---NSGKTPSIAAGVVGGLLVVGLGIGLYLRR-HIVKRTLRLLQEREL 113

QY 697 VEPLTPSGAMPNQAMRILKETELRKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLR 756
Db 114 VEPLTPSGEAPNQAHLRIKETEFKVKVLSGAFGVYKGIWIPDGENVKIPVAIKELR 173

QY 757 ENTSPKANKEILDAYMAGVSPVSRLLGICLTSTVQLTQLMYPYGLLDHVRNRRGR 816
Db 174 EATSPKANKEILDAYMNASVDNPHVCRLLGICLTSTVQLTQLMYPYGLLDYIREHKDN 233

QY 817 LGSODLLNMCQIAKMSYLEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETE 876
Db 234 IGSQVLLNWCQIAKMWYLERLVRDLAARNVVKTPQHVKITDFGLAKLGADEKEY 293

QY 877 YHADGKVPYIKWMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLE 936
Db 294 YHAEGKVPYIKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSVLE 353

QY 937 KGERLPQPPICTIDVYIMVKWMDSECRPRFELVSEFMRMDPQRFVVIQ-NEDLG 995
Db 354 KGERLPQPPICTIDVYIMVKWMDSDASRKPRFELIAEFKSMARDPPRYLVIQGDERRMH 413

QY 996 PASPLDSTFYRSLLDEDDMGDLVDAEYLVPPQGFCCPDAPGAGGMVHRHRSSTRSG 1055
Db 414 LPSTFDSKFYRLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454

QY 1056 GGDLTGLLEPSEERAPRSL-----APSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPQLQ 1110
Db 455 -----SRTPLLSLSATSNNATNCIDRNGG-----H----- 481

QY 1111 RYSEDPTVPLPSETDGVAPLTCSQPEYVNPQDVRPQPPSPREGPLPAARPAAGAT-LER 1169
Db 482 -----PVREDGFL-----PAPEYVNO--LMPKKPSTAMVQNIYVYISLTATSK 523

QY 1170 AKTLSPGKGVVQDVAFGGAVENPEYL 1197
Db 524 LPIDSRYN-----SHSTAVDNPAYL 544
```

RESULT 14

S00727

kinase-related transforming protein (erbB) (EC 2.7.1.1-) - avian erythroblastosis virus
C:Species: avian erythroblastosis virus

C>Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997

C:Accession: S00727

R.Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.

Oncogene Res. 1, 265-278, 1987

A:Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mutant

A:Reference number: S00727; MUID:88217326; PMID:2897102

A:Accession: S00727

A:Molecule type: DNA

A:Residues: 1-545 <SCO>

A:Cross-references: EMBL:X06943

C:Genetics:

A:Gene: erbB

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; phosphotransferase

F:135-400/Domain: protein kinase homology <KIN>

F:143-151/Region: protein kinase ATP-binding motif

```
Query Match      24.1%; Score 1640; DB 2; Length 545;
Best Local Similarity 54.9%; Pred. No. 2.9e-60;
Matches 345; Conservative 69; Mismatches 122; Indels 92; Gaps 15;

Qy 578 GPEADQCACAHYKDPFCVACPCSGVKKPDLSYMPIWKPFBEGACQPCPCINCTHSCVDL 637
Db 1 GP--DHCCKCAHFDGPHCVKACPAVLGENDTL-VKRYADANAVCOLCHPNCNCTRGCKGP 57
Qy 638 DDKGPACQASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETEL 696
Db 58 GLEGCP-----NGSKTPSIAGVVGGLCLLVVGLGIGLYLRR-HIVKRTLRLLQEREL 113
Qy 697 VEPLTPGAMPNQAMRILKETELRKVKVLGSGAFGTGVIWIPDGENVKIPVAIKVLR 756
Db 114 VEPLTPSGEAPNQAHRLILKETEFKKVKVLGFGAFGTGVIWIPGEKVTIPVAIKELR 173
Qy 757 ENTSPKANKEILDEAYVMAGVSPVSRLLGLCLTSTVOLVTQMLPYGCLLDHVRNRR 816
Db 174 EATSPKANKEILDEAYVMASVDNPHVCRLLGLCLTSTVOLITQMLPYGCLLDYIREHKN 233
Qy 817 LGSODLLNWCQIAKGMVLEEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 876
Db 234 IGSQYLLNWCQIAKGMVLEERHVLVHRDLAARNVLKTPQDKVITDFGLAKQLGADKE 293
Qy 877 YHADGGKVPICKMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLE 936
Db 294 YHAEKGKVPICKMALESILRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSVLE 353
Qy 937 KGERLPQPICTIDVYIMVWKWIDSECRPRRELVSBEFSRMARDPQRFVVIQ-NEDLG 995
Db 354 KGERLPQPICTIDVYIMVWKWMSDADSRPKRELIAEFKSMARDPPRYLVIOGDERMH 413
Qy 996 PASPLDSTFYRSLLEDMDGLVDAAEYLVPOQGFPCDPAPGAGGMVHHRSSSTRSG 1055
Db 414 LPSPTDSKFYRTLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
Qy 1056 GGDLTGLPESEEA PRSPL-----APSEGAGSDVFDGDLGMAAKGLQLSPLTHDPSPLQ 1110
Db 455 -----SRTPLLSLSLTSNNSATNCIDRNGG-----H----- 481
Qy 1111 RYSEDPTVLPSETDGYVAPLTCSPQPEYVNPQDVRPQPSREGPLPAARPAGAT-LER 1169
Db 482 -----PVREDGFL-----PAPEYVNQ--LMPKKPSTAMVQNIYNYISLTAISK 523
Qy 1170 AKTILSPGKGVVQVFAFGGAVENPEYL 1197
Db 524 LPMDSRYQN-----SHSTAVDNPEYL 544
```

RESULT 15

```
B44776
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain ES4)
C:Species: avian erythroblastosis virus
C:Date: 28-Apr-1993 #sequence_revision 28-Apr-1993 #text_change 04-Feb-2000
C:Accession: B44776
R:Bruskin, A.; Jackson, J.; Bishop, J.M.; McCarley, D.J.; Schatzman, R.C.
Oncogene 5, 15-24, 1990
A:Title: Six amino acids from the retroviral gene gag greatly enhance the transforming p
A:Reference number: A44776; PMID:90206603; PMID:1969616
A:Accession: B44776
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-540 <BRU>
A:Cross-references: GB:X52211
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; phosphotransferase; tyrosine-specific protein kinase
F:130-395/Domain: protein kinase homology <KIN>
F:138-146/Region: protein kinase ATP-binding motif
```

```
Query Match      23.9%; Score 1623; DB 2; Length 540;
Best Local Similarity 54.9%; Pred. No. 1.4e-59;
Matches 340; Conservative 69; Mismatches 120; Indels 90; Gaps 14;

Qy 587 CAHYKDPFCVACPCSGVKKPDLSYMPIWKPFBEGACQPCPCINCTHSCVDLDDKGCPEAQ 646
Db 3 CAHFDGPHCVKACPAVLGENDTL-VKRYADANAVCOLCHPNCNCTRGCKGGLGECGP--- 58
Qy 647 RASPLTSTIVSAVV-GILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGA 705
Db 59 NGSKTPSIAGVVGGLCLLVVGLGIGLYLRR-HIVKRTLRLLQERELVEPLTPSGE 117
Qy 706 MPNQAQMRILKETELRKVKVLGSGAFGTGVIWIPDGENVKIPVAIKVLRNTPSPKANK 765
Db 118 APNQAHLRIILKETEFKKVKVLGFGAFGTGVIWIPGEKVTIPVAIKELREATSPKANK 177
Qy 766 EILDEAYVMAGVSPVSRLLGLCLTSTVOLVTQMLPYGCLLDHVRNRRGLGSODLLNW 825
Db 178 EILDEAYVMASVDNPHVCRLLGLCLTSTVOLITQMLPYGCLLDYIREHKNIGSYLLNW 237
Qy 826 CMQIAKGMVLEEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEYHADGGKVP 885
Db 238 CVQIAKGMVLEERHVLVHRDLAARNVLKTPQDKVITDFGLAKQLGADKEKEYHAEKGKVP 297
Qy 886 IKWMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPP 945
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSVLEKGERLPQPP 357
Qy 946 ICTIDVYIMVWKWIDSECRPRRELVSBEFSRMARDPQRFVVIQ-NEDLG PASPLDSTF 1004
Db 358 ICTIDVYIMVWKWMSGADSRPKRELIAEFKSMARDPPRYLVIOGDERMHLPSPTDSKF 417
Qy 1005 YRSLEDDMDGLVDAAEYLVPOQGFPCDPAPGAGGMVHHRSSSTRSGGDLTLGLE 1064
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1065 PSEEA PRSPL-----APSEGAGSDVFDGDLGMAAKGLQLSPLTHDPSPLQRYSEDPTVP 1119
Db 450 -----SRTPLLSLSLTSNNSATNCIDRNGG-----H----- 476
Qy 1120 LPSETDGYVAPLTCSPQPEYVNPQDVRPQPSREGPLPAARPAGAT-LERAKTILSPGKN 1178
Db 477 -PVREDGFL-----PAPEYVNQ--LMPKKPSTAMVQNIYNYISLTAISKLPMDSRYN 527
Qy 1179 GWKVDKFAFGGAVENPEYL 1197
Db 528 -----SHSTAVDNPEYL 539
```

Search completed: July 22, 2003, 09:09:55

Job time : 30.9062 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 22.9062 Seconds
(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-579-593-12

Perfect score: 6803

Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 9613422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6651	97.8	1255	1 A24571	protein-tyrosine k
2	5856	86.1	1260	1 TVRTNU	protein-tyrosine k
3	5846.5	85.9	1254	2 I48161	p-185 precursor -
4	3116	45.8	1210	1 GQHUE	epidermal growth f
5	3089	45.4	1210	2 A53183	epidermal growth f
6	3064.5	45.0	1223	1 TVCHLV	epidermal growth f
7	2956.5	43.5	1308	2 A47253	epidermal growth f
8	2641	38.8	1166	1 S06142	protein-tyrosine k
9	2388.5	35.1	1342	2 A36223	kinase-related tra
10	2304.5	33.9	1339	2 J43387	epidermal growth f
11	1736.5	25.5	698	1 TVFVLV	protein-tyrosine k
12	1688	24.8	604	1 TVYUHV	protein-tyrosine k
13	1630.5	24.0	1330	1 G0PFE	epidermal growth f
14	1617	23.8	544	2 S35745	protein-tyrosine k
15	1610	23.7	545	2 S00727	kinase-related tra
16	1593	23.4	540	2 B44776	protein-tyrosine k
17	1591	23.4	540	1 TVFVEB	protein-tyrosine k
18	1477	21.7	644	2 A36325	epidermal growth f
19	1294	19.0	1323	2 E88257	protein let-23 (im
20	1294	19.0	1374	2 S70712	protein-tyrosine k
21	1211	17.8	1369	2 S70713	protein-tyrosine k
22	1176	17.3	1717	1 A45558	epidermal growth f
23	1126	16.6	527	2 A42032	epidermal growth f
24	976.5	14.4	843	2 A27131	epidermal growth f
25	806.5	11.9	346	2 S13807	protein-tyrosine k
26	754.5	11.1	311	2 S13808	protein-tyrosine k
27	723	10.6	1363	2 T43220	insulin-like growth
28	697	10.2	1607	2 T43212	insulin-like growth
29	691	10.2	1383	2 A36080	insulin receptor p

RESULT 1

A24571

protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N:Alternate names: C-erb-B-2 protein precursor; kinase-related transforming protein erbB

C:Species: Homo sapiens (man)

C>Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999

C:Accession: A24571; A25491; A44188; B44188; I59509; I57622

R:Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;

Nature 319, 230-234, 1986

A>Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth f

A:Reference number: A24571; MUID:86118663; PMID:3003577

A:Accession: A24571

A:Molecule type: mRNA

A:Residues: 1-1255 <YAM>

A:Cross-references: GB:X03363; NID:g31197; PIDN:CAA27060.1; PID:g31198

R:Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.

Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A>Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epiderm

A:Reference number: A25491; MUID:86016729; PMID:2995967

A:Accession: A25491

A:Molecule type: DNA

A:Residues: 737-1031 <SEM>

A:Cross-references: GB:M11767; NID:g182163; PIDN:AAA35808.1; PID:g553282

R:Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg, P

Science 230, 1132-1139, 1985

A>Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromos

A:Reference number: A44188; MUID:86070181; PMID:2999974

A:Accession: A44188

A:Molecule type: DNA

A:Residues: 740-910 <COUL>

A:Cross-references: GB:M12036; NID:g183988; PIDN:AAA35978.1; PID:g183989

A:Accession: B44188

A:Molecule type: mRNA

A:Residues: 1-517; RALL, 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>

A:Cross-references: GB:M11730; NID:g183986

R:King, C.R.; Kraus, M.H.; Aaronson, S.A.

Science 229, 974-976, 1985

A>Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A:Reference number: I59509; MUID:85272597; PMID:2992089

A:Accession: I59509

A>Status: translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 832-909 <REX>

A:Cross-references: GB:I29395; NID:g459807; PIDN:AAA35809.1; PID:g459808

R:Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987

A>Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptional

A:Reference number: I57622; MUID:87286898; PMID:3039351

A:Accession: I57622

A>Status: translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-191 <TAL>

A;Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
C;Comment: Amplification and overexpression of this erDB-related gene occurs in about 30
C;Genetics:
A;Gene: GDB:ERBB2; NGL; NEU; HER-2
A;Cross-references: GDB:120613; OMIM:164870
A;Map position: 17q21.1-17q21.1
A;Introns: 25/1; 75/3; 147/1; 883/3
A;Note: the list of introns is incomplete
C;Function:
A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
inase
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-125/Domain: protein-tyrosine kinase erBB2 #status predicted <MAT>
F:22-653/Domain: extracellular #status predicted <EXT>
F:70-304/Domain: EGF receptor extracellular domain repeat <EE1>
F:395-605/Domain: EGF receptor extracellular domain repeat <EE2>
F:654-675/Domain: transmembrane #status predicted <TM>
F:676-1255/Domain: intracellular #status predicted <INT>
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif
F:68.124,187,259,530,571.629/Binding site: carboxylate (Asn) (covalent) #status predicted
F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:753/Active site: Lys #status predicted
F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 97.8%; Score 6651; DB 1; Length 1255;
Best Local Similarity 97.6%; Pred. No. 2.1e-265;
Matches 1225; Conservative 10; Mismatches 20; Indels 0; Gaps 0;

```
QY 1 MELAALCWGLLLALPPCASTVCTGDMKRLPASPETHDMLRLHYGCGVQGNL 60
DB 1 MELAALCWGLLLALPPCASTVCTGDMKRLPASPETHDMLRLHYGCGVQGNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGVYLIHNNQVQLRQLRIVRGQLFEDNYVALVDNG 120
DB 61 ELTYLPTNASLSFLQDIOEVQGVYLIHNNQVQLRQLRIVRGQLFEDNYVALVDNG 120
QY 121 DPLNTTPTVGTASPGGLRELQRLSLEILKGVLIQRPOLQCYQDTILWKDIFHKNNOLA 180
DB 121 DPLNTTPTVGTASPGGLRELQRLSLEILKGVLIQRPOLQCYQDTILWKDIFHKNNOLA 180
QY 181 LTLIDNRSRACHPCSPCKSGRWGSESDCOSLTRVTCAGGCARCKGPLPTDCCHQC 240
DB 181 LTLIDNRSRACHPCSPCKSGRWGSESDCOSLTRVTCAGGCARCKGPLPTDCCHQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNLYSTDVGSCTLVCLNQNQVTDGTCRCEKSKPCARVCYGLGMQVYKANSKFIQIT 360
DB 301 YNLYSTDVGSCTLVCLNQNQVTDGTCRCEKSKPCARVCYGLGMQVYKANSKFIQIT 360
QY 361 ELEFAGCKKIFGSLAFLESPDGPASNTAPLOPEQLQVFEITGLVLYISAMPDSLP 420
DB 361 IOEFAGCKKIFGSLAFLESPDGPASNTAPLOPEQLQVFEITGLVLYISAMPDSLP 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGSLWGLRSLRELGSGLALIHNNTHLCFVHTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGSLWGLRSLRELGSGLALIHNNTHLCFVHTV 480
QY 481 PWDQLFRNPQALLTANRPEDECVGEGLAHQLCARGHCWGPQTQCVCNSQFLRGQEC 540
DB 481 PWDQLFRNPQALLTANRPEDECVGEGLAHQLCARGHCWGPQTQCVCNSQFLRGQEC 540
QY 541 VEECRVLQGLPREYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPECVARC 600
DB 541 VEECRVLQGLPREYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPECVARC 600
QY 601 PQYIKANSKFIQITELPDEEGACQPCINCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660
```

```
DB 601 PSGVKPDLSTYMPKWFPEEGACQPCINCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660
QY 661 ILLVVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
DB 661 ILLVVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
QY 721 RKVKVLGSGAGFTGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKELIDEAYVMAGVSP 780
DB 721 RKVKVLGSGAGFTGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKELIDEAYVMAGVSP 780
QY 781 YVSRLLGICLTSTVQLVTQMLPYCCLLDHVRNRRGLSQDLLNWCMIQAKGMSYLEDVR 840
DB 781 YVSRLLGICLTSTVQLVTQMLPYCCLLDHVRNRRGLSQDLLNWCMIQAKGMSYLEDVR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPDKWMALESILRRRFT 900
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPDKWMALESILRRRFT 900
QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQPPICITIDVYIMVKCWM 960
DB 901 HQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQPPICITIDVYIMVKCWM 960
QY 961 IDSECRPRFRELVSERFARMARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDDDDMGDLVDA 1020
DB 961 IDSECRPRFRELVSERFARMARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDDDDMGDLVDA 1020
QY 1021 BEYLVPOQGFPCPDPAAGAGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
DB 1021 BEYLVPOQGFPCPDPAAGAGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
QY 1081 AGSDVDFGDLGMAAGKLSLQTHDPSPLOKYSDDPTVPLPSETDGVVAPLTCSPQPEYV 1140
DB 1081 AGSDVDFGDLGMAAGKLSLQTHDPSPLOKYSDDPTVPLPSETDGVVAPLTCSPQPEYV 1140
QY 1141 NQPDVRPQPPSPREGPLPAARPGATLERAKTSLPGKNGVVKDYFAFGAVENPEYLTTPQ 1200
DB 1141 NQPDVRPQPPSPREGPLPAARPGATLERAKTSLPGKNGVVKDYFAFGAVENPEYLTTPQ 1200
QY 1201 GGAAPOHPHPAPSPAFDNLYYWDQPPERCAPPSTFGTPTAENPEYLGIDVPV 1255
DB 1201 GGAAPOHPHPAPSPAFDNLYYWDQPPERCAPPSTFGTPTAENPEYLGIDVPV 1255
```

RESULT 2
TVRTNU
protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
C;Accession: A24562; A61204
R;Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A;Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.
A;Reference number: A24562; MUID:86118662; PMID:3945311
A;Accession: A24562
A;Molecule type: mRNA
A;Residues: 1-1260 <BAR>
A;Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746
R;Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen,
Carcinogenesis 12, 1975-1978, 1991
A;Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no m
2-thiazolyl]formamide or N-methyl-N-nitrosourea.
A;Reference number: A61204; MUID:92035293; PMID:1682063
A;Accession: A61204
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 637-663,'V',665-702 <MAS>
A;Note: authors translated the codon GCA for residue 25 as Val
C;Genetics:
A;Gene: neu
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

F:658-680/Domain: transmembrane #status predicted <TN>
F:723-988/Domain: protein kinase homology <KIN>
F:731-739/Region: protein kinase ATP-binding motif
F:711,191,263,535,576,634/Binding site: carboxydrate (Asn) (covalent) #status predicted
F:691/Binding site: phosphate (Thr) (covalent) #status predicted
F:758/Active site: Lys #status predicted
F:1882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

```
Query Match      86.1%; Score 5856; DB 1; Length 1260;
Best Local Similarity 86.0%; Pred. No. 7, 9e-233;
Matches 1081; Conservative 56; Mismatches 118; Indels 2; Gaps 2;

Qy 1 MELAALCRWGLLALLPPGAASVCTGTDMKLRPLASPETHDMLRLHYGCGVQVQNL 60
Db 4 MELAALCRWGLLALLPPGIAAGTCTGTDMKLRPLASPETHDMLRLHYGCGVQVQNL 63
Qy 61 ELTYLPTNASLFLQDIQEVGYVLIHNOVROVPLQRLIRVRGTLFEDNYALAVLDNG 120
Db 64 ELTYVPAVASLFLQDIQEVGYVLIHNOVROVPLQRLIRVRGTLFEDNYALAVLDNR 123
Qy 121 DPLNNTTPTV-CASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILKDIHKNQNL 179
Db 124 DPQDNVAASVTPGTEPEGLRELQRLSLTEILKGGVLIQRPOLCYQDMVLAKDVRKNNQL 183
Qy 180 ALTLIDNRSRACHPCSPMKGSRGWGSSSDCQSLTRTVAGGCARCKGRLPTDCCHEQ 239
Db 184 APVDIDNRSRACHPCSPMKGSRGWGSSSDCQSLTRTVAGGCARCKGRLPTDCCHEQ 243
Qy 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTAC 299
Db 244 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTTC 303
Qy 300 PNYLSTDVGSCTLVCPLNNOEVTAEADGTQRCCKSKPCARVCYGLGMOYTKANSKFTGI 359
Db 304 PNYLSTEVGSCTLVCPLNNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLRGARITSD 363
Qy 360 TELEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLOVFETLEETIGYLYISAWPDSL 419
Db 364 NVQFDFGCKKIFGSLAFIPESFDGDPSSGIAPLRPEQLQVFETLEETIGYLYISAWPDSL 423
Qy 420 PDLVSFONLQVIRGRILHNGAYSITLQGLGISWGLRSLRELGSGLALIHNTLHLCFVHT 479
Db 424 RDLVSFONLQVIRGRILHNGAYSITLQGLGISWGLRSLRELGSGLALIHNTLHLCFVHT 483
Qy 480 VPWDLFRNPQALLHTANRPEDE-CVGEGLACHOLCARGCWGPGTQCVCNCSQFLRGQ 538
Db 484 VPWDLFRNPQALLHSGNRPEEDLCVSSGLVCSNLAHGCWGPQTQCVCNCSHFLRGQ 543
Qy 539 ECVBEICRVLOGLPREYVNAHCLPCHPECQPNQSVTCFGEADQCAAHYKDPFFCVA 598
Db 544 ECVBEICRVWGLPREYVSDKRLPCHPECQPNQSVTCFGEADQCAAHYKDSSCVA 603
Qy 599 RCPQYIKANSKFIGITELPDEGACOPCINCTHSCVDLDDKGPAPORASPLTSIVAV 658
Db 604 RCPGVKPDLSYMPITWKYPDEBEGICQPCINCTHSCVDLDDKGPAPORASPLTSIVAV 663
Qy 659 VGILLVAVLVGVFGLIKRRQKIRKYTMRLLOTELVEPLTPSGAMPNOAQRILKET 718
Db 664 EGVLLFLILVVVGLIKRRQKIRKYTMRLLOTELVEPLTPSGAMPNOAQRILKET 723
Qy 719 ELRKVKVLGSAFGTVYKGIWIPGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVG 778
Db 724 ELRKVKVLGSAFGTVYKGIWIPGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVG 783
Qy 779 SPYVSRLLIGICLTSTVQLVTOIMPYGCLLDHVRNRLGSGODLLNWCQIAKMSYLE 838
Db 784 SPYVSRLLIGICLTSTVQLVTOIMPYGCLLDHVRNRLGSGODLLNWCQIAKMSYLE 843
Qy 839 VRLVHRLDAARNVLKSPNHVKITDFGLARLLDDETEYHADGGKVPKIKWALSIILRRR 898
Db 844 VRLVHRLDAARNVLKSPNHVKITDFGLARLLDDETEYHADGGKVPKIKWALSIILRRR 903
Qy 899 FTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMIWVC 958
```

```
Db 904 FTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMIWVC 963
Qy 959 WMIDSECRPRFRELVSFSESMARDPQRFVVIQNEIDLCPASPLDSTFVRSLEDDMDGLV 1018
Db 964 WMIDSECRPRFRELVSFSESMARDPQRFVVIQNEIDLCPASPLDSTFVRSLEDDMDGLV 1023
Qy 1019 DAEYLVPQOGFFCPCDPAPGAGVMVHRHRSSTRSGGDLTLGLEPSEEBAPRSLAPS 1078
Db 1024 DAEYLVPQOGFFCPCDPAPGAGVMVHRHRSSTRSGGDLTLGLEPSEEBAPRSLAPS 1083
Qy 1079 EGAGSDVFDGDLGMAAKGLOSPLTHDPSPFLORYSEDPVPLPSETDGYVAPLTCPOPE 1138
Db 1084 EGAGSDVFDGDLGMAAGVTKQLSLPHLSPLQRYSEDPVPLPSETDGYVAPLTCPOPE 1143
Qy 1139 YVNOQDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPGKNGVWVDVAFGAVENPEYLT 1198
Db 1144 YVNOQDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPGKNGVWVDVAFGAVENPEYLV 1203
Qy 1199 PQGGAAPQPPHPPAFSPAFDNLVYWDQDPPERGAAPPSTFKGTPTAENPEYLGLDVVP 1255
Db 1204 PREGTASPPHSPAFSPAFDNLVYWDQDPPERGAAPPSTFKGTPTAENPEYLGLDVVP 1260

RESULT 3
148161
p-185 precursor - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C:Accession: I48161
R:Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa,
Gene 140, 251-255, 1994
A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A:Reference number: I48161; MUID:94193007; PMID:7908275
A:Accession: I48161
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1254 <RES>
A:Cross-references: GB:D16295; NID:G493236; PID:BAA03801.1; PID:G747595
C:Genetics:
A:Gene: neu
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif

Query Match      85.9%; Score 5846.5; DB 2; Length 1254;
Best Local Similarity 85.7%; Pred. No. 1, 9e-232;
Matches 1075; Conservative 65; Mismatches 114; Indels 1; Gaps 1;

Qy 1 MELAALCRWGLLALLPPGAASVCTGTDMKLRPLASPETHDMLRLHYGCGVQVQNL 60
Db 1 MELAALCRWGLLALLPPGASGTVCTGTDMKLRPLASPETHDMLRLHYGCGVQVQNL 60
Qy 61 ELTYLPTNASLFLQDIQEVGYVLIHNOVROVPLQRLIRVRGTLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLFLQDIQEVGYVLIHNOVROVPLQRLIRVRGTLFEDNYALAVLDNR 120
Qy 121 DPLNNTTPTV-CASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILKDIHKNQNL 180
Db 121 DPLNNTTPTV-CASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILKDIHKNQNL 180
Qy 181 LTLDITNRSRACHPCSPMKGSRGWGSSSDCQSLTRTVAGGCARCKGRLPTDCCHEQ 240
Db 181 PVDIDNRSRACHPCSPMKGSRGWGSSSDCQSLTRTVAGGCARCKGRLPTDCCHEQ 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTAC 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTTC 300
Qy 301 YNYLSTDVGSCTLVCPLNNOEVTAEADGTQRCCKSKPCARVCYGLGMOYTKANSKFTGI 360
Db 301 YNYLSTEVGSCTLVCPLNNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLRGARITSAN 360
```


A: Molecule type: mRNA
A: Residues: 969-971, 'K', 973-1115, 'D' <EIS>
A: Cross-references: EMBL:Z12608
R: Heiser, G. J.; Gill, G. N.
J. Biol. Chem. 263, 13152-13158, 1988
A: Title: Epidermal growth factor receptor threonine and serine residues phosphorylated in
A: Reference number: A28941; MUID: 88330814; PMID: 3138233
A: Accession: A28941
A: Molecule type: protein
A: Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000, 1002-1009,
R: Hibbs, M. L.; Dunn, A. R.; Alexander, W. S.
submitted to the EMBL Data Library, April 1994
A: Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
A: Reference number: S45325
A: Accession: S45325
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-971, 'K', 973-1210 <VER>
A: Cross-references: EMBL: X78987; NID: G488830; PID: CAAS5587.1; PID: G488831
R: Paria, B. C.; Das, S. K.; Andrews, G. K.; Dey, S. K.
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
A: Title: Expression of the epidermal growth factor receptor gene is regulated in mouse b
A: Reference number: I49643; MUID: 93126380; PMID: 7678348
A: Accession: I49643
A: Status: translated from GB/EMBL/DBJ
A: Molecule type: mRNA
A: Residues: 12-20, 22-132 <RES>
A: Cross-references: GB: L06864; NID: G193001; PID: AAA53029.1; PID: G567201
C: Genetics:
A: Gene: EGFR
C: Superfamily: epidermal growth factor receptor; protein kinase homology
C: Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphoprotein
F: 1-24/Domain: signal sequence #status predicted <SIG>
F: 648-670/Domain: transmembrane #status predicted <TM>
F: 712-977/Domain: protein kinase homology <KIN>
F: 720-728/Region: protein kinase ATP-binding motif
F: 680, 695/Binding site: phosphate (Thr) (covalent) #status experimental
F: 697, 1070/Binding site: phosphate (Ser) (covalent) #status experimental
F: 993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
F: 1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
F: 1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 45.4%; Score 3089; DB 2; Length 1210;
Best Local Similarity 49.6%; Pred. No. 1.6e-119;
Matches 632; Conservative 165; Mismatches 364; Indels 112; Gaps 25;

Qy 11 LLLALLPGCAA--STQVCTGTDMLRLPASPTHLMDLRLHLYQGVQGVQGNLELYLPTN 68
Db 14 LTLALCAAGGALLEKKVCGQTSNRLTQGTDFHFLSLQRYNCEVVLNLEITYVQRN 73
Qy 69 ASLSFLQDIQVQGYVLIHANOVQVQLRILVRGTQLFEDNYALAVLDNGDPLNNTTP 128
Db 74 YDLSEFKTIQVAGVYVLIANTVRIPLNLQIRGNALYENTYALAILSN----- 124
Qy 129 VTGASPGGLRELQRLSLTEILKGVLIQBNPOLCYQDTILWDI----FHKNNQLALTLI 184
Db 125 -YGTNRTGLRELPMNLQELIGAVRFNSNNPILCNNDTIQWRDVIQNVFMNMSMDL--- 180
Qy 185 DNNRSRACHPCSPMKGRCWSESSEDCOSLRTVCAGGCA-RCKGLPPTDCCHQCAAG 243
Db 181 -QSHFSSCPKCDPSPNCSWGGGNCQKLTKIACQCSHRCGRSPSDCHNQCAAG 239
Qy 244 CTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPNY 303
Db 240 CTGPRESDCLVCQKQDEATCKDTCPLMLYNPTTYQMDVNPPEGKYSFGATCVKCKPRNY 299
Qy 304 LSTDVGSCTLVCLPHNQVTAEDGTQRCCKSPCARVCYGLGMQYIKANSKFIGITILE 363
Db 300 VYTDHGSCVRACGPDYVEY-EDGIRKCKKCDGPCRKVCNGIGIEFK-DTLISINATNIK 357
Qy 364 -FAGCKKIFGSLAFIPESFDGPPASNTAPLQEQVFTLEITGYLYISAWPDSLPDL 422
Db 358 HFKYCTAISGDHILPLVAPKGSFTTTPPLDPRELEILTKVETITGFLLIQAPDNWIDL 417

RESULT 6

TVCHLV

epidermal growth factor receptor precursor - chicken

N: Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB

C: Species: Gallus gallus (chicken)

C: Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000

C: Accession: A27720; A00643

R: Iax, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennet

Mol. Cell. Biol. 8, 1970-1978, 1988

A: Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou

A: Reference number: A27720; MUID: 88261272; PMID: 3260329

A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
A:Cross-references: GB:M20386
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro-
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: A00643
A:Molecule type: mRNA
A:Residues: 595-1223 <NIL>
A:Cross-references: GB:M10066
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
specific protein kinase
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F:31-654/Domain: extracellular #status predicted <EXT>
F:81-307/Domain: EGF receptor extracellular domain repeat <EB1>
F:397-610/Domain: EGF receptor extracellular domain repeat <EB2>
F:655-677/Domain: transmembrane #status predicted <TM>
F:678-1223/Domain: intracellular #status predicted <INT>
F:719-984/Domain: protein kinase homology <KIN>
F:727-735/Region: protein kinase ATP-binding motif
F:136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #
F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:754/Active site: Lys #status predicted
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #sta-

QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRFT 900
 Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRFT 900
 QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYMWKCMW 960
 Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYMWKCMW 960
 QY 961 IDSECRPRRELVSERFARMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDMDGLVDA 1020
 Db 961 IDSECRPRRELVSERFARMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDMDGLVDA 1020
 QY 1021 EYVLVPOQGFCDPAPACAGVHHRSSSTRSGGDLTLGLEPSEEEAPRSLPSEG 1080
 Db 1021 EYVLVPOQGFCDPAPACAGVHHRSSSTRSGGDLTLGLEPSEEEAPRSLPSEG 1080
 QY 1081 AGSDVFDGDLGMAAGLQSLPHTDPSLQRYSEDPTVPLPSETDGVVAPLTCSPOPEYV 1140
 Db 1081 AGSDVFDGDLGMAAGLQSLPHTDPSLQRYSEDPTVPLPSETDGVVAPLTCSPOPEYV 1140
 QY 1141 NQPDVRRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAPGGAVENPEYLTPQ 1200
 Db 1141 NQPDVRRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAPGGAVENPEYLTPQ 1200
 QY 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
 Db 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 6

AAW92406
 ID AAW92406 standard; Protein; 1255 AA.

AC AAW92406;
 XX
 XX 21-APR-1999 (first entry)
 DT
 XX Human HER-2/neu oncogene protein.
 DE
 XX HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
 KW malignancy; treatment; tumour.
 KM
 XX Homo sapiens.

XX Key Location/Qualifiers
 FH 676..1255
 FT Region /note= "region which elicits immune response"
 FT
 XX US5869445-A.
 PN
 XX 09-FEB-1999.
 XX
 XX 01-APR-1996; 96US-0625101.
 XX
 XX 01-APR-1996; 96US-0625101.
 PR 17-MAR-1993; 93US-0033644.
 PR 12-AUG-1993; 93US-0106112.
 PR 31-MAR-1995; 95US-0414417.
 XX
 XX (UNIW) UNIV WASHINGTON.

XX Cheever MA, Disis ML;
 XX WPI; 1999-152835/13.
 DR N-PSDB; AAX01912.
 DR

XX Use of HER-2/neu polypeptides - for eliciting an immune response to
 PT an HER-2/neu associated malignancy, particularly for treating or
 PT preventing tumours
 XX
 PS Claim 3; Column 31-38; 26pp; English.

XX This sequence represents the human HER-2/neu oncogene protein. A fragment

CC of this protein is used in a method for eliciting or enhancing an immune
 CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
 CC B cells to produce an immune response to the HER-2/neu protein. The
 CC method can be used for immunisation against a malignancy in which the
 CC HER-2/neu oncogene is associated and in the treatment of an existing
 CC tumour, or to prevent tumour occurrence or reoccurrence.

XX Sequence 1255 AA;

Query Match 99.9%; Score 6806; DB 20; Length 1255;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1253; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MELAALCRNGLLLALLPPGAASSTQVCTGTMKRLRPPASPEHLDMRLHLYQGQVVOGNL 60

Db 1 MELAALCRNGLLLALLPPGAASSTQVCTGTMKRLRPPASPEHLDMRLHLYQGQVVOGNL 60

QY 61 ELYLPTNASLSFLQDIOEVQGVLIHNNQVPLQRLRIRVGTOLFEDNYALAVLDNG 120

Db 61 ELYLPTNASLSFLQDIOEVQGVLIHNNQVPLQRLRIRVGTOLFEDNYALAVLDNG 120

QY 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQORNQVPLQVQDTILWKDIFHKNNQLA 180

Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQORNQVPLQVQDTILWKDIFHKNNQLA 180

QY 181 LTLIDTNRSRACHPCSPMCKGSRMGESSEDCQSLTRTVACGACARCKGPLPTDCCHEQC 240

Db 181 LTLIDTNRSRACHPCSPMCKGSRMGESSEDCQSLTRTVACGACARCKGPLPTDCCHEQC 240

QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP 300

Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP 300

QY 301 YNYLSTDVGSCTLVCPLNHNOEVTAEQTCRCKSKPCARVCYGLGMEHLREVAVTSAN 360

Db 301 YNYLSTDVGSCTLVCPLNHNOEVTAEQTCRCKSKPCARVCYGLGMEHLREVAVTSAN 360

QY 361 IOEFAGCKKIFGSLAPLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLISAWPDSL 420

Db 361 IOEFAGCKKIFGSLAPLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLISAWPDSL 420

QY 421 DLSVFONLQVIRGRILHNGAYSILTLQGLIGISWGLRSLRELGLALIHNNTHLCFVHTV 480

Db 421 DLSVFONLQVIRGRILHNGAYSILTLQGLIGISWGLRSLRELGLALIHNNTHLCFVHTV 480

QY 481 PNDQLFRNPHQALLHTANRPEDECVGEGACQPCINCTHSCVDLDDKGCAPAEQASPLTSI 540

Db 481 PNDQLFRNPHQALLHTANRPEDECVGEGACQPCINCTHSCVDLDDKGCAPAEQASPLTSI 540

QY 541 VEECRVLOGLPREYVYNARHCLFCHPECPQNGSVTCFGEADQCVACAHYKDPFFCVARC 600

Db 541 VEECRVLOGLPREYVYNARHCLFCHPECPQNGSVTCFGEADQCVACAHYKDPFFCVARC 600

QY 601 PSGVKPDLISYMPIWKPFPDEGACQPCINCTHSCVDLDDKGCAPAEQASPLTSI 660

Db 601 PSGVKPDLISYMPIWKPFPDEGACQPCINCTHSCVDLDDKGCAPAEQASPLTSI 660

QY 661 ILLVVLGVVFGILIKRROOKIRKYTMRELLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Db 661 ILLVVLGVVFGILIKRROOKIRKYTMRELLQETELVEPLTPSGAMPNQAQMRILKETEL 720

QY 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYMAGVGSF 780

Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYMAGVGSF 780

QY 781 YVSRLLGICLTSTVQLVTOLMPYGLLDHVNRNRLGSGQDLLNWCMIKAGMSYLEYDR 840

Db 781 YVSRLLGICLTSTVQLVTOLMPYGLLDHVNRNRLGSGQDLLNWCMIKAGMSYLEYDR 840

QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRFT 900

Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRFT 900

QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPPICITIDVYIMVYKWM 960
DB 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPPICITIDVYIMVYKWM 960
QY 961 IDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDMDGLDVA 1020
DB 961 IDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDMDGLDVA 1020
QY 1021 BEYLVPQGFPCPDPAAGAGVHHRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
DB 1021 BEYLVPQGFPCPDPAAGAGVHHRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
DB 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTISPGKNGVVKOVFAFGGAVENPEYLTPQ 1200
DB 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTISPGKNGVVKOVFAFGGAVENPEYLTPQ 1200
QY 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
DB 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 7

ID AAB21198 standard; protein; 1255 AA.
AC AAB21198;
XX
DT 12-JAN-2001 (first entry)
XX Human HER-2/neu protein.
XX Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
XX colon cancer.
OS Homo sapiens.
XX
PN WO200044899-A1.
XX
XX 03-AUG-2000.
XX
PF 28-JAN-2000; 2000WO-US02164.
XX
PR 29-JAN-1999; 99US-0117976.
XX (CORI-) CORIXA CORP.
PA (SMIK) SMITHKLINE BEECHAM.
XX
PI Cheever MA, Gheysen D;
XX WPI; 2000-505976/45.
DR N-PSDB; AAA89736.
XX
XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT useful for vaccinating against breast, ovarian, colon, lung and
PT prostate cancers -
XX
XX Claim 52; Fig 7; 128pp; English.
XX
CC The present sequence is the human HER-2/neu protein. It is a member of
CC the tyrosine kinase family of receptor-like glycoproteins and shows
CC homology to the epidermal growth factor receptor (EGFR). It probably
CC plays a part in cell growth and/or differentiation. The HER-2/neu
CC gene is an oncogene. An HER-2/neu fusion protein comprising a
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC domain may be used to treat or prevent cancer by eliciting or
CC enhancing an immune response to the HER-2/neu protein. It may be used
CC to treat malignancies such as breast, ovarian, colon, lung and
CC prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.
XX
SQ Sequence 1255 AA;
Query Match 99.9%; Score 6806; DB 21; Length 1255;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1253; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MELAALCRGLLALLPCCAASQVCTGDMKLRLPASPETHLMLRHLVQGCQVQGNL 60
DB 1 MELAALCRGLLALLPCCAASQVCTGDMKLRLPASPETHLMLRHLVQGCQVQGNL 60
QY 61 EUTYLPNASLSFLQDIQEVQGVLIHNRQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
DB 61 EUTYLPNASLSFLQDIQEVQGVLIHNRQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
QY 121 DPLNNTPTVGTASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIPHKQNLA 180
DB 121 DPLNNTPTVGTASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIPHKQNLA 180
QY 181 LTLIDTNRSRACHPCSPMCKGRCWGESSEDCOSLTRTVCCAGCARCKGPLPTDCCHEQC 240
DB 181 LTLIDTNRSRACHPCSPMCKGRCWGESSEDCOSLTRTVCCAGCARCKGPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACILHFNHSGICEHLCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACILHFNHSGICEHLCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHNOEVTAEQTCRCKSKPCARVCYGLGMEHLREVRAVTSAN 360
DB 301 YNYLSTDVGSCTLVCPHNOEVTAEQTCRCKSKPCARVCYGLGMEHLREVRAVTSAN 360
QY 361 IOEFAGCKKIFGSLAPLPSFDGDPASNTAPLQPEOLQVFETLEETGYLTSAPDLSLP 420
DB 361 IOEFAGCKKIFGSLAPLPSFDGDPASNTAPLQPEOLQVFETLEETGYLTSAPDLSLP 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWLGRLSRLSGLALIHNTLHLCFVHTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWLGRLSRLSGLALIHNTLHLCFVHTV 480
QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGACHQLCARGHCWPGPTQCVCNCSQFLRGQEC 540
DB 481 PWDQLFRNPHQALLHTANRPEDECVGEGACHQLCARGHCWPGPTQCVCNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFVCARC 600
DB 541 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFVCARC 600
QY 601 PSGVKPDLSPYMPIWKPFPDEEGACQPCPINCTHSCVDLDDKGCPEAQASPLTSIVSAVVG 660
DB 601 PSGVKPDLSPYMPIWKPFPDEEGACQPCPINCTHSCVDLDDKGCPEAQASPLTSIVSAVVG 660
QY 661 ILLVVVLGVVFGILIKRROQKIRKVTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
DB 661 ILLVVVLGVVFGILIKRROQKIRKVTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 721 RKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYYVAGVGP 780
DB 721 RKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYYVAGVGP 780
QY 781 YVSRLLIGICLTSTVQLVTLQMPYGCILLDHVRENRRGLSGQDLLNMCWIAKMSYLEDVYR 840
DB 781 YVSRLLIGICLTSTVQLVTLQMPYGCILLDHVRENRRGLSGQDLLNMCWIAKMSYLEDVYR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWALLESILRRRFT 900
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWALLESILRRRFT 900
QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPPICITIDVYIMVYKWM 960
DB 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPPICITIDVYIMVYKWM 960
QY 961 IDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDMDGLDVA 1020

Db 961 IDSECRFRRELVSERFMRDPRFVVIQNEEDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
 QY 1021 EYLVPQGGFFCPDPAPGAGMHHRRSSSTRSGGDLTLGLEPSEEEAPRSPAPSEG 1080
 Db 1021 EYLVPQGGFFCPDPAPGAGMHHRRSSSTRSGGDLTLGLEPSEEEAPRSPAPSEG 1080
 QY 1081 AGSDVFGDILGMAAGLQSLPTHPDSPLOQRYSEDPVPLPSETDGVVAPLTCSPQPEYV 1140
 Db 1081 AGSDVFGDILGMAAGLQSLPTHPDSPLOQRYSEDPVPLPSETDGVVAPLTCSPQPEYV 1140
 QY 1141 NOPDVRQPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVAFGGAVENPEYLTPQ 1200
 Db 1141 NOPDVRQPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVAFGGAVENPEYLTPQ 1200
 QY 1201 GGAAPQHPHPAPSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
 Db 1201 GGAAPQHPHPAPSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 8

AAY84780

ID AAY84780 standard; Protein; 1255 AA.

XX AAY84780;

08-AUG-2000 (first entry)

XX Amino acid sequence of the SPLICE erbB-2 receptor protein.

XX SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;
 KW tumor cell proliferation; tissue degeneration; arthropathy;
 KW bone resorption; inflammatory disease; degenerative disorder;
 KW wound healing.

XX Homo sapiens.

XX W0200020579-A1.

XX 13-APR-2000.

XX 01-OCT-1999; 99WO-CA00912.

XX 02-OCT-1998; 98US-0165192.

XX (UYMC-) UNIV MCMASTER.

XX Muller WJ, Siegel PM;

XX WPI; 2000-303768/26.

XX N-PSDB; AAA14812.

XX Nucleic acid encoding an erbB 2 receptor protein designated SPLICE
 PT erbB-2, inhibitors of the protein are useful for treatment of cancer -
 XX Claim 3; Fig 2; 60pp; English.

XX The present sequence represents a SPLICE erbB-2 receptor protein. The
 CC protein has an in-frame deletion of 16 amino acids, 2 of which are
 CC conserved cysteine residues, compared to the unspliced protein. The
 CC erbB-2 polynucleotide is used to construct probes for detecting
 CC disorders of cell transformation such as cancer. Antibodies to the
 CC protein may be used to detect SPLICE erbB-2 in a sample. Agents
 CC (e.g. antisense oligonucleotides) which inhibit the expression of
 CC SPLICE erbB-2 are useful for reducing tumor cell proliferation and
 CC treating cancer. Substances which stimulate SPLICE erbB-2 are useful
 CC for treating conditions involving damaged cells including conditions
 CC in which degeneration of tissue occurs, such as arthropathy, bone
 CC resorption, inflammatory diseases, degenerative disorders of the
 CC central nervous system and wound healing.

XX Sequence 1255 AA;

Query Match 99.9%; Score 6806; DB 21; Length 1255;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1253; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MELAAALCRWGLLLALLPAGAASTQVCTGTDMLRLPASPTHLDMLRHLYQGCVVQGNL 60
 Db 1 MELAAALCRWGLLLALLPAGAASTQVCTGTDMLRLPASPTHLDMLRHLYQGCVVQGNL 60
 QY 61 ELTYLPTNASLFLQDIQEVQGVLIHAHQVQLQRLRIVRGTQLFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLFLQDIQEVQGVLIHAHQVQLQRLRIVRGTQLFEDNYALAVLDNG 120
 QY 121 DELNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFPHKNQLA 180
 Db 121 DELNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFPHKNQLA 180
 QY 181 LTLIDTNRGRACHPCSPCKGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCCHEQC 240
 Db 181 LTLIDTNRGRACHPCSPCKGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCCHEQC 240
 QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 QY 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRAVTSAN 360
 Db 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRAVTSAN 360
 QY 361 IOEFACKKIFGSLAPLPSFDGDPASNTAPLOPELOVFTLEETITGVLYISAWPDSL 420
 Db 361 IOEFACKKIFGSLAPLPSFDGDPASNTAPLOPELOVFTLEETITGVLYISAWPDSL 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGSIWLGRLSLRELGLALIHNTLHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGSIWLGRLSLRELGLALIHNTLHLCFVHTV 480
 QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGACHQLCARGHCWGPPTQCVNCSQFLRGQEC 540
 Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGACHQLCARGHCWGPPTQCVNCSQFLRGQEC 540
 QY 541 VBECEVLQGLFREYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKPPFCVARC 600
 Db 541 VBECEVLQGLFREYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKPPFCVARC 600
 QY 601 PSGVKPDLSPYMPITWKPEDEGACQPCPINCTHSCVDLDDKGCPEAORASPLTSIVSAVVG 660
 Db 601 PSGVKPDLSPYMPITWKPEDEGACQPCPINCTHSCVDLDDKGCPEAORASPLTSIVSAVVG 660
 QY 661 ILLVVVLGVVFGILIKRROQKIRKVTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
 Db 661 ILLVVVLGVVFGILIKRROQKIRKVTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
 QY 721 RKVKVLGSGAFGTVYKGIWIPGENVKIPVAIKVLRNTSPKANKILDEAYMAGVGPSP 780
 Db 721 RKVKVLGSGAFGTVYKGIWIPGENVKIPVAIKVLRNTSPKANKILDEAYMAGVGPSP 780
 QY 781 YVSRLLGICLTSTVOLVTQLMYPVGLLDHVRNRRGLSGQDLLNWCQIAKGSYLEVDYR 840
 Db 781 YVSRLLGICLTSTVOLVTQLMYPVGLLDHVRNRRGLSGQDLLNWCQIAKGSYLEVDYR 840
 QY 841 LVHRDLAARNVLKGNPNHVKITDFGLARLLDIDETEHADGGKVPKMWALESIILRRFT 900
 Db 841 LVHRDLAARNVLKGNPNHVKITDFGLARLLDIDETEHADGGKVPKMWALESIILRRFT 900
 QY 901 HQSDVWSYGVYTWELMTFGAKYDGIIPAREIPDLLEKGERLPPOPICTTIDVTMVKWCM 960
 Db 901 HQSDVWSYGVYTWELMTFGAKYDGIIPAREIPDLLEKGERLPPOPICTTIDVTMVKWCM 960
 QY 961 IDSECRPRELVSERFMRDPRFVVIQNEEDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
 Db 961 IDSECRPRELVSERFMRDPRFVVIQNEEDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
 QY 1021 EYLVPQGGFFCPDPAPGAGMHHRRSSSTRSGGDLTLGLEPSEEEAPRSPAPSEG 1080

Db	1021	EEYLVPOQGFPCDPAPAGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG	1080
QY	1081	AGSDVFDGDLGMAAGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV	1140
Db	1081	AGSDVFDGDLGMAAGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV	1140
QY	1141	NQDVRPQPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVAFGGAVENPEYLTPO	1200
Db	1141	NQDVRPQPSPREGPLPAARPAAGATLERPKTILSPGKNGVYKDVAFGGAVENPEYLTPO	1200
QY	1201	GGAAPQHPHPPAFSPFDNLVYWDQDPPRGAPSTFKGTPTAENPEYLGLDVVP	1255
Db	1201	GGAAPQHPHPPAFSPFDNLVYWDQDPPRGAPSTFKGTPTAENPEYLGLDVVP	1255
RESULT 9			
AAB85458			
ID	AAB85458 standard; Protein; 1255 AA.		
XX	AAB85458;		
XX	25-SEP-2001 (first entry)		
XX	Human HER-2/neu protein.		
XX	Antigen-presenting cell; immunogenic; immune response; HER-2/neu;		
KW	oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.		
XX	Homo sapiens.		
XX	WO200153463-A2.		
XX	26-JUL-2001.		
XX	19-JAN-2001; 2001WO-US01850.		
XX	21-JAN-2000; 2000US-0177545.		
XX	(CORI-) CORIXA CORP.		
XX	Cheever MA, Hand-Zimmermann S;		
XX	WPI; 2001-476112/51.		
DR	N-PSDB; AAH23392.		
XX	New antigen-presenting cells, useful as vaccines for eliciting or		
PT	enhancing an immune response to HER-2/neu protein, particularly useful		
PT	for treating or preventing cancer, e.g. breast cancer -		
XX	Claim 2; Page 41-46; 49pp; English.		
XX	The invention provides an isolated antigen-presenting cell, which		
CC	expresses at least an immunogenic portion of a polypeptide that produces		
CC	an immune response to HER-2/neu protein. The antigen-presenting cells are		
CC	useful as vaccines for eliciting or enhancing an immune response to		
CC	HER-2/neu protein, particularly in treating or preventing malignancies in		
CC	which the HER-2/neu oncogene is associated. Specifically, these are		
CC	useful for treating or preventing cancer, e.g. breast cancer, ovarian,		
CC	colon, lung or prostate cancers. The present sequence represents		
CC	the human HER-2/neu protein (also known as p185 or c-erbB2).		
XX	SQ Sequence 1255 AA;		
Query Match 99.9%; Score 6806; DB 22; Length 1255;			
Best Local Similarity 99.8%; Pred. No. 0;			
Matches 1253; Conservative 1; Mismatches 1; Indels 0; Gaps 0;			
QY	1	MELAALCRWGLLALLPPGAASQVCTGDMKLRLPASPTHLDMLRHLVQGCQVQGNL	60
Db	1	MELAALCRWGLLALLPPGAASQVCTGDMKLRLPASPTHLDMLRHLVQGCQVQGNL	60
QY	61	ELTYLPTNASLSFLQDIQEVGYVLIHAHQVRQVPLQRLIRVGTQLFEDNYALAVLDNG	120

Db	61	ELTYLPTNASLSFLQDIQEVGYVLIHAHQVRQVPLQRLIRVGTQLFEDNYALAVLDNG	120
Qy	121	DPLNNTPTVPGASPGGLRELQRLSLEILKGGVLIQORNQOLCYQDTILKNDIIFHNKOLA	180
Db	121	DPLNNTPTVPGASPGGLRELQRLSLEILKGGVLIQORNQOLCYQDTILKNDIIFHNKOLA	180
Qy	181	LTLIDNTRSRACHPCSPCKGRCWGESSEDCOSLTRTVCAGGCARCKGKPLPTDCCHEQC	240
Db	181	LTLIDNTRSRACHPCSPCKGRCWGESSEDCOSLTRTVCAGGCARCKGKPLPTDCCHEQC	240
Qy	241	AAGCTGPKHSDCLACILHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP	300
Db	241	AAGCTGPKHSDCLACILHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP	300
Qy	301	YNLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRATISAN	360
Db	301	YNLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRATISAN	360
Qy	361	IOEFAGCKKIFGSLAFPLPESFDGDPASNTAPLQPEQLQVFETLEETGYLYISAWPDSL	420
Db	361	IOEFAGCKKIFGSLAFPLPESFDGDPASNTAPLQPEQLQVFETLEETGYLYISAWPDSL	420
Qy	421	DLSVFQNLQVIRGRILHNGAYSILTLQGLIGISWLGRLSRLSGSLALIHNTLHLCFVHTV	480
Db	421	DLSVFQNLQVIRGRILHNGAYSILTLQGLIGISWLGRLSRLSGSLALIHNTLHLCFVHTV	480
Qy	481	PMDQLFRNPHQALLHTANRPEDECVGEGLAGLQCHARGCWGPGTQCVCNSQFLRGQEC	540
Db	481	PMDQLFRNPHQALLHTANRPEDECVGEGLAGLQCHARGCWGPGTQCVCNSQFLRGQEC	540
Qy	541	VBECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFVCVARC	600
Db	541	VBECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFVCVARC	600
Qy	601	PSGKPDLSYMPITWPKPDEEGACQPCPINCTHSCVDLDDKCPAEQASPLTSIVSAVVG	660
Db	601	PSGKPDLSYMPITWPKPDEEGACQPCPINCTHSCVDLDDKCPAEQASPLTSIVSAVVG	660
Qy	661	ILLVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAOMRILKETEL	720
Db	661	ILLVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAOMRILKETEL	720
Qy	721	RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKIILDEAYVMAGVSP	780
Db	721	RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKIILDEAYVMAGVSP	780
Qy	781	YVSRLLGICLTSTVQLVTLQMPYGLLDHVRENRLGSLQDLNWCMIKAGMSYLEDVR	840
Db	781	YVSRLLGICLTSTVQLVTLQMPYGLLDHVRENRLGSLQDLNWCMIKAGMSYLEDVR	840
Qy	841	LVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALLESILRRFT	900
Db	841	LVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALLESILRRFT	900
Qy	901	HOSDVMVSGYVWELMTFGAKPVGDIIPAREIPDLLEKGERLPPOPCTIDVYMIWKCM	960
Db	901	HOSDVMVSGYVWELMTFGAKPVGDIIPAREIPDLLEKGERLPPOPCTIDVYMIWKCM	960
Qy	961	IDSECRPRFRELVSFSESRMARDPQRFVVIQNEEDLGPASPLDSTFYRSLLDDMDGLVDA	1020
Db	961	IDSECRPRFRELVSFSESRMARDPQRFVVIQNEEDLGPASPLDSTFYRSLLDDMDGLVDA	1020
Qy	1021	BEYLVPQOGFFCPDDPAPGAGMVHRRSSSTRSGGDLTLGLEPSEEAAPRSLADSEG	1080
Db	1021	BEYLVPQOGFFCPDDPAPGAGMVHRRSSSTRSGGDLTLGLEPSEEAAPRSLADSEG	1080
Qy	1081	AGSDVFDGLGMGAAGLQSLPTHDPSPLORYSEDPTVPLPSETDGVVAPLTCSPQPEYV	1140
Db	1081	AGSDVFDGLGMGAAGLQSLPTHDPSPLORYSEDPTVPLPSETDGVVAPLTCSPQPEYV	1140
Qy	1141	NQDVRQPQPSREGPLPAARPAAGATLERAKTILSPGKNGVVVDVAFGGAVENPEYLTPO	1200

Db	1141	NQPDVVRQPPSPREGFLPAARPAAGATLERPKTLSPGKNGVVDVFAFGAVENPEVLTPO	1200
QY	1201	GGAAPOPHPPAPAFDNLNLYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVP	1255
Db	1201	GGAAPOPHPPAPAFDNLNLYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVP	1255
RESULT 10			
AAG88267			
ID	AAG88267	standard; Protein; 1255 AA.	
XX	AAG88267;		
XX	11-SEP-2001	(first entry)	
XX	HER2/neu	amino acid sequence.	
XX	Human; HER2/neu;	epitope; human leukocyte antigen; HLA; T cell;	
KW	immune response; vaccine;	tumour; cancer; cytostatic; immunostimulant;	
KW	tumour-associated antigen; T lymphocyte;	cytotoxic T lymphocyte; CTL.	
XX	Homo sapiens.		
XX	WO200141787-A1.		
PD	14-JUN-2001.		
XX	11-DEC-2000;	2000WO-US33591.	
XX	10-DEC-1999;	99US-0458299.	
XX	(EPIM-) EPIMUNE INC.		
PI	Pikes J, Sette A, Sidney J, Southwood S, Cheenut R, Celis E;		
PI	Keogh E;		
XX	WPI; 2001-374995/39.		
DR	An isolated prepared HER2/neu epitope useful in a vaccine for inducing		
PT	cellular immune responses for the prevention and treatment of cancer -		
PT	Disclosure; Page 15; 199pp; English.		
PS	The present invention describes isolated prepared HER2/neu epitopes (I).		
XX	Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is		
CC	culture in vitro and binds to a complex of an epitope (I), bound to a		
CC	human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)		
CC	and a second epitope and the peptide is less than 50 contiguous amino		
CC	acids that have 100% identity with a native peptide sequence of HER2/neu;		
CC	(3) a vaccine composition (III) comprising (II) and a pharmaceutical		
CC	excipient; (4) an isolated nucleic acid encoding a peptide comprising		
CC	(I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic		
CC	and immunostimulant activities, and can be used in vaccines. (I), (II)		
CC	prevention and treatment of cancer. (I) and (II) are useful for		
CC	monitoring or evaluating an immune response to a tumour-associated		
CC	antigen when incubated with a T lymphocyte sample from a patient and		
CC	detecting the presence of bound T lymphocyte to (I) or (II). Epitope		
CC	based vaccines mean that immunosuppressive epitopes that may be present		
CC	in whole antigens may be avoided. Selected epitopes may be combined to		
CC	enhance immunogenicity. The possible pathological side effects caused by		
CC	infectious agents or whole protein antigen is eliminated. The vaccine		
CC	provides the ability to direct and focus an immune response to multiple		
CC	selected antigens from the same pathogen. Epitope-based anti-tumour		
CC	vaccines provides the opportunity to combine epitopes derived from		
CC	multiple tumour-associated molecules addressing the problem of tumour-		
CC	tumour variability and reducing the likelihood of tumour escape due to		
CC	antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in		
CC	the exemplification of the present invention.		
XX	Sequence	1255 AA;	
SQ	Query Match	99.9%; Score 6806; DB 22; Length 1255;	

Db 1021 EYLVPPQGFPCDPAPGAGGMVHRHRSSTRSGGDLTLGLBPSEEEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPQPPSPREGPLPAARPAGATLERAKTSLPGKNGVVKDYVAFGGAVENPEYLTPQ 1200
Db 1141 NOPDVRPQPPSPREGPLPAARPAGATLERPKTSLFGKNGVVKDYVAFGGAVENPEYLTPQ 1200
Qy 1201 GGAAPQHPHPPAFSPADNLYYWDODPPERCAPSTFKGTPTAENPEYGLDVPV 1255
Db 1201 GGAAPQHPHPPAFSPADNLYYWDODPPERCAPSTFKGTPTAENPEYGLDVPV 1255

RESULT 11

AAE24067
ID AAE24067 standard; Protein; 1255 AA.

XX AC
AC AAE24067;

XX 23-SEP-2002 (first entry)

XX Human Her-2 protein.

XX Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KW hyperproliferative disorder; prophylaxis; inflammation; antisense;
KW tumour; gene therapy; phosphorothioate backbone.

XX Homo sapiens.

XX WO200222636-A1.

XX 21-MAR-2002.

XX 12-SEP-2001; 2001WO-US28572.

XX 15-SEP-2000; 2000US-0663834.

XX (ISIS-) ISIS PHARM INC.

XX Bennett CF, Cowse LM;

XX WPI; 2002-471192/50.

XX N-PSDB; AAD38904.

XX Novel antisense oligonucleotide which modulates the expression of Human
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT inflammation or to prevent infection in humans -

XX Example 13; Page 95-107; 116pp; English.

XX The invention relates to antisense compounds targetted to a nucleic
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC that specifically hybridises with and inhibits the expression of Her2.
CC Antisense compounds of the invention are used for treating diseases or
CC conditions associated with Her2 such as hyperproliferative disorders
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC neural or cardiac cancer. They are also useful prophylactically e.g.
CC to prevent or delay infection, inflammation and tumour formation. The
CC invention is also used in gene therapy. The present sequence is human
CC Her-2 protein.

XX Sequence 1255 AA;

Query Match 99.9%; Score 6806; DB 23; Length 1255;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1253; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLALLPPGAASQVCTGTDKMLRLPASPEHLDMLRHLVQGCQVQGNL 60
Db 1 MELAALCRWGLLALLPPGAASQVCTGTDKMLRLPASPEHLDMLRHLVQGCQVQGNL 60

Qy 61 ELTYLPTNASLSLQDIQEVQGVYLAHNQVRQVPLQRLRIVRGTQOLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSLQDIQEVQGVYLAHNQVRQVPLQRLRIVRGTQOLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVGASPGGLREQLRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHNKQLA 180
Db 121 DPLNNTTPTVGASPGGLREQLRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHNKQLA 180
Qy 181 LTLIDNTRSPACHPCSPMCKGSCWGESSEDCSLTRTVCCAGCARCKGPLPTDCCHEQC 240
Db 181 LTLIDNTRSPACHPCSPMCKGSCWGESSEDCSLTRTVCCAGCARCKGPLPTDCCHEQC 240
Qy 241 AAGCTGPKGSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTACP 300
Db 241 AAGCTGPKGSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTACP 300
Qy 301 YNYLSTDVSGCTLVCPHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTACP 360
Db 301 YNYLSTDVSGCTLVCPHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTACP 360
Qy 361 IQEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPEOLQVPELLEEITGYLYISAWPDSL 420
Db 361 IQEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPEOLQVPELLEEITGYLYISAWPDSL 420
Qy 421 DLSVFQNLQVIRGRIHNGAYSLTLOGLISWLGRLSLRELGSGLALIHNTLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRIHNGAYSLTLOGLISWLGRLSLRELGSGLALIHNTLHLCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGTQCVNCSQFLRGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGTQCVNCSQFLRGQEC 540
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECQONGSVTCFGEADOCVACAHYKDPFCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECQONGSVTCFGEADOCVACAHYKDPFCVARC 600
Qy 601 PSGVKPDLSPYMPKPFDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSISAVVG 660
Db 601 PSGVKPDLSPYMPKPFDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSISAVVG 660
Qy 661 ILLVVVLGVVFGILIKRROQKIRKYTMRRLLOTELVEPLTPSGAMPNOAQRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRROQKIRKYTMRRLLOTELVEPLTPSGAMPNOAQRILKETEL 720
Qy 721 RKVKVLGSGAFGTVYKGIWIPGENVKIIPVALKVLRENTSPKANKILDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGTVYKGIWIPGENVKIIPVALKVLRENTSPKANKILDEAYVMAGVGP 780
Qy 781 YVSRLLGICLTSTVQLVTQMPYGCILLDHVRENRLGSLQDLINMCWQIAKMSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTQMPYGCILLDHVRENRLGSLQDLINMCWQIAKMSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALBSILRRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALBSILRRRFT 900
Qy 901 HOSDWSYGVYVWELMTFCAKPDGIPAREIPDLLEKGERLPQPPCTIDVYMWKCMW 960
Db 901 HOSDWSYGVYVWELMTFCAKPDGIPAREIPDLLEKGERLPQPPCTIDVYMWKCMW 960
Qy 961 IDSECRPRELVSEFSRMARDPQRFVITQNEEDLGPASPLDSTFYRSLLDDDDMGDLVDA 1020
Db 961 IDSECRPRELVSEFSRMARDPQRFVITQNEEDLGPASPLDSTFYRSLLDDDDMGDLVDA 1020
Qy 1021 EYLVPPQGFPCDPAPGAGGMVHRHRSSTRSGGDLTLGLBPSEEEAPRSLAPSEG 1080
Db 1021 EYLVPPQGFPCDPAPGAGGMVHRHRSSTRSGGDLTLGLBPSEEEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPQPPSPREGPLPAARPAGATLERAKTSLPGKNGVVKDYVAFGGAVENPEYLTPQ 1200

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 22.9062 Seconds
(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-369-383-12
Perfect score: 6815
Sequence: 1 MELAALCRWGLLLALLPPCA.....TFKGTPTAENPEYLGIDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73:*

1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6653	97.6	1255	1 A24571	protein-tyrosine k
2	5859	86.0	1260	1 TVRTNU	protein-tyrosine k
3	5851.5	85.9	1254	2 I48161	p-185 precursor -
4	3097	45.4	1210	1 GQHUE	epidermal growth f
5	3089	45.0	1210	2 A53183	epidermal growth f
6	3052.5	44.8	1223	1 TVCHLV	epidermal growth f
7	2919.5	42.8	1308	2 A47253	epidermal growth f
8	2636	38.7	1166	1 S06142	protein-tyrosine k
9	2366.5	34.7	1342	2 A36223	kinase-related tra
10	2280.5	33.5	1339	2 JC4387	epidermal growth f
11	1766.5	25.9	698	1 TVFVLV	protein-tyrosine k
12	1703	25.0	604	1 TVYUHV	protein-tyrosine k
13	1647	24.2	544	2 S35745	protein-tyrosine k
14	1640	24.1	545	2 S00727	protein-tyrosine k
15	1623	23.8	540	2 B44776	kinase-related tra
16	1621	23.8	540	1 TVFVEB	protein-tyrosine k
17	1600.5	23.5	1330	1 GQFE	protein-tyrosine k
18	1461	21.4	644	2 A36325	epidermal growth f
19	1275	18.7	1323	2 E88257	epidermal growth f
20	1275	18.7	1374	2 S70712	protein let-23 [im
21	1192	17.5	1369	2 S70713	protein-tyrosine k
22	1161	17.0	1717	1 A45558	epidermal growth f
23	1084	15.9	527	2 A42032	epidermal growth f
24	945.5	13.9	843	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	epidermal growth f
26	754.5	11.1	311	2 S13808	protein-tyrosine k
27	730	10.7	1363	2 T43220	protein-tyrosine k
28	697	10.2	1382	1 INHUR	insulin-like growt
29	696	10.2	1383	2 A36080	insulin receptor p

30	696	10.2	1607	2 T43212	insulin-like growt
31	695.5	10.2	1372	2 A34157	insulin receptor p
32	676	9.9	1477	2 T18534	protein-tyrosine k
33	673.5	9.9	1300	2 A36502	insulin receptor-r
34	664	9.7	1268	2 B36502	insulin receptor-r
35	644	9.4	1367	1 IGHUR1	insulin-like growt
36	627	9.2	1390	2 T30346	insulin receptor -
37	625	9.2	1371	2 A33837	insulin-like growt
38	615.5	9.0	2148	1 A56081	insulin receptor (
39	612	9.0	2101	2 S57245	insulin receptor (
40	599	8.8	987	2 A54092	protein-tyrosine k
41	588	8.6	1114	1 S05582	protein-tyrosine k
42	587.5	8.6	984	2 A39753	protein-tyrosine k
43	586	8.6	1091	2 S33596	protein-tyrosine k
44	585.5	8.6	952	2 I50612	protein-tyrosine k
45	581.5	8.5	977	2 S49004	tyrosine kinase Mp

ALIGNMENTS

RESULT 1

A24571
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
N:Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein erbB2
C:Species: Homo sapiens (man)
C:Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999
C:Accession: A24571; A25491; A44188; B44188; I59509; I57622
R:Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;
Nature 319, 230-234, 1986
A:Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth factor
A:Reference number: A24571; MUID:86118663; PMID:3003577
A:Accession: A24571
A:Molecule type: mRNA
A:Residues: 1-1255 <YAM>
A:Cross-References: GB:X03363; NID:g31197; PIDN:CAA27060.1; PID:g31198
R:Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A:Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epidermal growth factor
A:Reference number: A25491; MUID:86016729; PMID:2995967
A:Accession: A25491
A:Molecule type: DNA
A:Residues: 737-1031 <SEM>
A:Cross-References: GB:M11767; NID:g182163; PIDN:AAA35808.1; PID:g553282
R:Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg, P.
Science 230, 1132-1139, 1985
A:Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal
A:Reference number: A44188; MUID:86070181; PMID:2999974
A:Accession: A44188
A:Molecule type: DNA
A:Residues: 740-910 <COU1>
A:Cross-References: GB:M12036; NID:g183988; PIDN:AAA35978.1; PID:g183989
A:Accession: B44188
A:Molecule type: mRNA
A:Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>
A:Cross-References: GB:M11730; NID:g183986
R:King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 229, 974-976, 1985
A:Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
A:Reference number: I59509; MUID:85272597; PMID:2992089
A:Accession: I59509
A:Molecule type: DNA
A:Status: translated from GB/EMBL/DBJ
A:Residues: 832-909 <REX>
A:Cross-References: GB:L29395; NID:g459807; PIDN:AAA35809.1; PID:g459808
R:Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987
A:Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptional
A:Reference number: I57622; MUID:87286898; PMID:3039351
A:Accession: I57622
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-191 <TAL>

A;Cross-references: GB:M16792; NID:G183983; PIDN:AAA58637.1; PID:G553332
C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C;Genetics:
A;Gene: GDB:ERBB2; NGL; NEU; HER-2
A;Cross-references: GDB:120613; OMIM:164870
A;Map position: 17q21.1-17q21.1
A;Introns: 25/1; 75/3; 147/1; 883/3
A;Note: the list of introns is incomplete
C;Function:
A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
inase
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F;22-653/Domain: extracellular #status predicted <EXT>
F;70-304/Domain: EGF receptor extracellular domain repeat <EE1>
F;395-605/Domain: EGF receptor extracellular domain repeat <EE2>
F;654-675/Domain: transmembrane #status predicted <TM>
F;676-1255/Domain: intracellular #status predicted <INT>
F;718-983/Domain: protein kinase homology <KIN>
F;726-734/Region: protein kinase ATP-binding motif
F;68,124,187,259,530,571,629/Binding site: carboxydrate (Asn) (covalent) #status predict
F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;753/Active site: Lys #status predicted
F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 97.6%; Score 6653; DB 1; Length 1255;
Best Local Similarity 97.6%; Pred. No. 6.5e-265;
Matches 1225; Conservative 8; Mismatches 22; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALLPPGAASQVCTGTDMLRLPASPEHLDMRLHYQGCVQGNL 60
DB 1 MELAALCRWGLLALLPPGAASQVCTGTDMLRLPASPEHLDMRLHYQGCVQGNL 60
QY 61 ELTYLPTNASLFLQDIQEVQGVLIHAHQVROVPLQRLIRVGTOLFFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLFLQDIQEVQGVLIHAHQVROVPLQRLIRVGTOLFFEDNYALAVLDNG 120
QY 121 DPLNNTPTVTGSPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFPKKNOLA 180
DB 121 DPLNNTPTVTGSPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFPKKNOLA 180
QY 181 LTLIDNRSRACHPCSPMCKSGCWGESSEDCQSLRTVCAGGCARCKGPLETDCCHQC 240
DB 181 LTLIDNRSRACHPCSPMCKSGCWGESSEDCQSLRTVCAGGCARCKGPLETDCCHQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVSGCTLVCPHINQEVTAEDGTORCEKSPCARVCYGLGMQVIKANSKFIGHT 360
DB 301 YNYLSTDVSGCTLVCPHINQEVTAEDGTORCEKSPCARVCYGLGMQVIKANSKFIGHT 360
QY 361 ELFEAGCKIFGSLAFIPESFDGDPASNTAPQVIKANSKFIGHTELTGVLITYSAWPDLSL 420
DB 361 IQBFAGCKIFGSLAFIPESFDGDPASNTAPLOEQVETUEETIGVLIYSAWPDLSL 420
QY 421 DLSVFQNLQVIRGIRLHNGAYSITLQGLGISWGLSLRELGSGLALIHNNHLCFVHTV 480
DB 421 DLSVFQNLQVIRGIRLHNGAYSITLQGLGISWGLSLRELGSGLALIHNNHLCFVHTV 480
QY 481 PDQLFRNPHQALLHTANRPEDSCVGEGLACHOLCARGHCWGPGPTQVCNCSOFLRGQEC 540
DB 481 PDQLFRNPHQALLHTANRPEDSCVGEGLACHOLCARGHCWGPGPTQVCNCSOFLRGQEC 540
QY 541 VEECRVLQGLPREYVVARHCLPCHPECPONGSVTCFGEADOCVCAHYKDPFCVCARC 600
DB 541 VEECRVLQGLPREYVVARHCLPCHPECPONGSVTCFGEADOCVCAHYKDPFCVCARC 600
QY 601 PSGVKPDLSTYMPFIWKPDEEGACQPCPINCTHSCVDLDDKGCPEAQSRASPLTSIVSAVVG 660

RESULT 2

TVRTNU

protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999

C:Accession: A24562; A61204

R:Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.

Nature 319, 226-230, 1986

A:Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.

A:Reference number: A24562; MUID:86118662; PMID:3945311

A:Accession: A24562

A:Molecule type: mRNA

A:Residues: 1-1260 <BAR>

A:Cross-references: EMBL:X03362; NID:G56745; PIDN:CAA27059.1; PID:G56746

R:Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen,

Carcinogenesis 12, 1975-1978, 1991

A:Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no

2-thiazolylformamide or N-methyl-N-nitrosourea.

A:Reference number: A61204; MUID:92035293; PMID:1682063

A:Accession: A61204

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 637-663, 'V', 665-702 <MAS>

A>Note: authors translated the codon GCA for residue 25 as Val

C:Genetics:

A:Gene: neu

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

F:658-680/Domain: transmembrane #status predicted <TMN>
F:723-988/Domain: protein kinase homology <KIN>
F:731-739/Region: protein kinase ATP-binding motif
F:71,191,263,535,576,634/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:691/Binding site: phosphate (Thr) (covalent) #status predicted
F:758/Active site: Lys #status predicted
F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 86.0%; Score 5859; DB 1; Length 1260;
Best Local Similarity 86.0%; Pred. No. 1.8e-232;
Matches 1081; Conservative 54; Mismatches 120; Indels 2; Gaps 2;

QY 1 MELAALCRWGLLALLPPGAASSTVCTGDMKRLRASPETHLDMLRHLYQGCVVQGNL 60
DB 4 MELAAWCRWGLLALLPPGIAGTQVCTGDMKRLRASPETHLDMLRHLYQGCVVQGNL 63
QY 61 ELTYLPTNASLSFLQDIQEVQGVLLIAHNVQVPLQRLIRVRGTQLFEDNVVALAVLDNG 120
DB 64 ELTYVPANASLSFLQDIQEVQGVLLIAHNVQVPLQRLIRVRGTQLFEDKVALAVLDNR 123
QY 121 DPLNNTPTVT-GASPGGLRELQRLSLEILKGGVLIQRPOLCYQDTILWKDIFHKNQL 179
DB 124 DPQDNVAASTPGRTPEGLRELQRLSLEILKGGVLIQRPOLCYQDMVLWKDVFKNQL 183
QY 180 ALTLIDTNSRACHPCSPMKSGRCSWGESSEDCQSLTRIVCAGGCARCKGPLPTDCCHQ 239
DB 184 APVIDTNSRACPPCAPACKNHCWGESPEDCQILTGTICTSGCACRCKGRPLPTDCCHQ 243
QY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAC 299
DB 244 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTTC 303
QY 300 PYNLSTVGSCTLVCPFLHNQVETAEADGTQRCCKSKPCARVCYGLGMOYIKANSFVIGI 359
DB 304 PYNLSTVGSCTLVCPFPNNQVETAEADGTQRCCKSKPCARVCYGLGMEHLRGARITSD 363
QY 360 TELEFAGCKIKIGSLAFIPESPDGDPASNTAPQYIKANSKFIGITELTGLYIISAPWDSL 419
DB 364 NVQEFQCKIKIGSLAFIPESPDGDPSSGIAPLRPEQLQVFFLEITGLYIISAPWDSL 423
QY 420 PDLVSFQNLQVIRGLTHNGAYSLTQGLIGSLWGLRSRELGSGLALTHNTHLCFVHT 479
DB 424 RDLVSFQNLQVIRGLTHNGAYSLTQGLIGSLWGLRSRELGSGLALTHNTHLCFVHT 483
QY 480 VPWQDLFRNPQALLHTANRPEDE-CVGBGLACHQLCARGHCWGPQTQVNCSCQFLRQ 538
DB 484 VPWQDLFRNPQALLHSGNRPEEDLCVSSGLVCNLSCAHGHGWGPQTQVNCSHFLRQ 543
QY 539 ECVEECRVLOGLPREVYNARHCLPCHPECPQNGSVTCFGEPAQCVAACHYKDPFFCVA 598
DB 544 ECVEECRVWKLPREYVSDKRLPCHPECPQNSSETCFGEADQCAACHYKDSSCVA 603
QY 599 RCPGSKVPDLVYMPIWKFDEGACOPCINCTHSCVDLDDKGCPAEQRASPLTSTVSAY 658
DB 604 RCPGSKVPDLVYMPIWKYDDEGICOPCINCTHSCVDLDERCPAEQRASPVTFIATV 663
QY 659 VGIILLVVLGVWFGIILKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKET 718
DB 664 EGVLFLFLVVLVVGIIKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKET 723
QY 719 ELRKVKVLGSGAPGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVAVG 778
DB 724 ELRKVKVLGSGAPGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVAVG 783
QY 779 SPVSVLLGICLTSTVOLVTLMPYGLLDHVRNREGRIGSQDILLNCKWQIAGNSYLED 838
DB 784 SPVSVLLGICLTSTVOLVTLMPYGLLDHVRNREGRIGSQDILLNCKWQIAGNSYLED 843
QY 839 VRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGKVPKIMWALESTLRR 898
DB 844 VRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGKVPKIMWALESTLRR 903
QY 899 FTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKC 958

DB 904 FTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKC 963
QY 959 WMIDSECRPRFRELVSERSMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDGMDLV 1018
DB 964 WMIDSECRPRFRELVSERSMARDPQRFVVIQNEDLGPSSPMDSTFYRSLLEDGMDLV 1023
QY 1019 DAEYLYVQOQ3FFCPDPAPAGAGMVHHRSSSTSGGGDLTLGLEPSEEEAPRPLAPS 1078
DB 1024 DAEYLYVQOQ3FFSPDPTPGTGSTAHRHRSSSTSGGGELTLGLEPSEEGPRPLAPS 1083
QY 1079 EGAGSDVDGDLGMAAGKLOSLPDPQRYSEDPVPLPSTGTVVAPLTCSPQPE 1138
DB 1084 EGAGSDVDGDLGAMGVTKGLOSLPDPQRYSEDPVPLPSTGTVVAPLTCSPQPE 1143
QY 1139 YYNQDVPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYLT 1198
DB 1144 YYNQSEVQPPPLTPEGPLPVPRPAGATLERKTLSPGKNGVVKDVFAGGAVENPEYLV 1203
QY 1199 POGGAAPQHPHPPAFSPADNLYYWDQPPPERGAPSTFKGPTTAENPEYLGLDVVP 1255
DB 1204 PREGTASPPHSPAFSPADNLYYWDQNSSEGGPPSPNEGTPTAENPEYLGLDVVP 1260

RESULT 3
I48161
p-185 precursor - golden hamster
C:Species: Mesocricetus auratus (Golden hamster)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C:Accession: I48161
R:Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa, Gene 140, 251-255, 1994
A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A:Reference number: I48161; MUID:94193007; PMID:7908275
A:Accession: I48161
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1254 <RES>
A:Cross-references: GB:D16295; NID:g493236; PIDN:BAA03801.1; PID:g747595
C:Genetics:
C:Gene: neu
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif

Query Match 85.9%; Score 5851.5; DB 2; Length 1254;
Best Local Similarity 85.7%; Pred. No. 3.7e-232;
Matches 1075; Conservative 64; Mismatches 115; Indels 1; Gaps 1;

QY 1 MELAALCRWGLLALLPPGAASSTVCTGDMKRLRASPETHLDMLRHLYQGCVVQGNL 60
DB 1 MELAAWCRWGLLALLSPGASGTQVCTGDMKRLRASPETHLDIVRHLYQGCVVQGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQGVLLIAHNVQVPLQRLIRVRGTQLFEDNVVALAVLDNG 120
DB 61 ELTYLPANATLSFLQDIQEVQGVLLIAHNVQVPLQRLIRVRGTQLFEDKVALAVLDNR 120
QY 121 DPLNNTPTVT-GASPGGLRELQRLSLEILKGGVLIQRPOLCYQDTILWKDIFHKNQLA 180
DB 121 DPLDNVTATGTPEGLRELQRLSLEILKGGVLIQRPOLCYQDTILWKDIFHKNQLA 180
QY 181 LTLIDTNSRACHPCSPMKSGRCSWGESSEDCQSLTRIVCAGGCARCKGPLPTDCCHQ 240
DB 181 PVIDITNSRACPPCAPACKNHCWGESPEDCQILTGTIAPRAVPAARARLPTDCCHQ 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAC 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTTC 300
QY 301 YNLYSTVGSCTLVCPFLHNQVETAEADGTQRCCKSKPCARVCYGLGMOYIKANSFVIGI 360
DB 301 YNLYSTVGSCTLVCPFLHNQVETAEADGTQRCCKSKPCARVCYGLGMEHLRGARITSA 360

A;Molecule type: mRNA
 A;Residues: 969-971,'K',973-1115,'D' <EIS>
 A;Cross-references: EMBL:Z12608
 R;Heisermann, G.J.; Gill, G.N.
 J. Biol. Chem. 263, 13152-13158, 1988
 A;Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
 A;Reference number: A28941; MUID:88330814; PMID:3138233
 A;Accession: A28941
 A;Molecule type: protein
 A;Residues: 689-694,'X',696-704,'L',706-707;989-992,'XX',995-996,'X',998-1000;1002-1009,
 R;Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
 submitted to the EMBL Data Library, April 1994
 A;Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
 A;Reference number: S45325
 A;Accession: S45325
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-971,'K',973-1210 <VER>
 A;Cross-references: EMBL:X78987; NID:g488830; PIDN:CAA55587.1; PID:g488831
 R;Paria, B.C.; Das, S.K.; Andrews, G.K.; Dev, S.K.
 Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
 A;Title: Expression of the epidermal growth factor receptor gene is regulated in mouse b
 A;Reference number: I49643; MUID:93126380; PMID:7678348
 A;Accession: I49643
 A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 12-20,22-132 <RES>
 A;Cross-references: GB:L06864; NID:g193001; PIDN:AAA53029.1; PID:g567201
 A;Genetics:
 C;Superfamily: epidermal growth factor receptor; protein kinase homology
 C;Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop
 F;1-24/Domain: signal sequence #status predicted <SIG>
 F;648-670/Domain: transmembrane #status predicted <TM>
 F;712-977/Domain: protein kinase homology <KIN>
 F;720-728/Region: protein kinase ATP-binding motif
 F;680,695/Binding site: phosphate (Thr) (covalent) #status experimental
 F;997,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
 F;993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
 F;1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
 F;1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 45.0%; Score 3069; DB 2; Length 1210;
 Best Local Similarity 49.4%; Pred. No. 1.9e-118;
 Matches 629; Conservative 165; Mismatches 367; Indels 112; Gaps 25;
 QY 11 LILALLPPGAA--STQVCTCTDMKRLPASPETHDMLRLHYOGCQVQVQGNLETLVLP 69
 DB 14 LTLALCAAGALBEKKVQCQTSNRLTQLGTFFDHLFLSLQRMYNVCEVVLGNLEITYVQRN 73
 QY 69 ASLSFLQIQEVOGYVLIHNOVROVPLQRLRIVRGTLQFEDNYALAVLDNGDPLNNTTP 128
 DB 74 YDLSFLKTIQVAGYVLIANTVERIPLENLQIIRGNALYENTYALILSN----- 124
 QY 129 VTGASPGGLRELQRLSRLTEILKGVLTQIRNPOLCYODTILWKDI----FHKNQLALTLI 184
 DB 125 -YGTNRITGLRELPMRLNQLIELIAGVRFNSNPILCNMDTIQWRDIVQNVFMSNMSMDL--- 180
 QY 185 DTRSRACHPCSPMKGSRGWGSESSDCQSLRTVACGGA-CKGPLPTDCHEQCAAG 243
 DB 181 -QSHPSCKPCDPCSPGSCGWGSEENCQKLTIIICAQCSHRCSRSPSDCHNQCAAG 239
 QY 244 CTGPKHSIDCLALHFNHSGICELHCPALVTYNTDTFESMENPEGRYTFGASCVTACPVNY 303
 DB 240 CTGPRSDCLVCKQFQDEATCKDTCPPLMYNTTITQMDVNPBGKISFGATCVKKCPRNY 299
 QY 304 LSTDVGSCTLVCPHLNHQVTAEDGTORCEKCKPCARVCYGLGMQYIKANSKFIGITELE 363
 DB 300 VVTDHSGSVACGPDYVEV-EEDGIRKCKKCDGCRKVCNGIGIGBEFK-DTILSINATNIK 357
 QY 364 -FAGCKIFGSLAFLESFDCGDPASNTAPQYIKANSKFIGITELTGLYIISAWPDSLPL 422
 DB 358 FFKYCTAISGDLHLPLVAFKGSDFTRTPPLDPRELEILTKVKEITGFLLIQAWPDNWTDL 417

RESULT 6
TUCHLV

epidermal growth factor receptor precursor - chicken

N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB

C;Species: Gallus gallus (chicken)

C;Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000

C;Accession: A27720; A00643

R;Lax, I.; Johnson, A.; Hawk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennet

Mol. Cell. Biol. 8, 1970-1978, 1988

A;Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mo

A;Reference number: A27720; MUID:88261272; PMID:3260329

QY 423 SVFQNLQVIRGRILHNGAYSLTLOGLIGISWLGRLSIRLGLALIHHTHNLCHFVHTVPW 482
 DB 418 HAFENLEIRGRITKQHGQFSLAVGLNITSLGRLSKLSISDGDVIISGNRLCYANTINW 477
 QY 483 DOLFNRPHQALHTANRPEDECVGEGLAHQICARGHCGWPGPTOCVNCQSLRQGEQVE 542
 DB 478 KKLFGTGNQTKIMNRAEKCKAVNVHVCNPLCSSEGCGWGPEDPCVSCQNVSRGECVE 537
 QY 543 ECRVLQGLPREVYNARHCLPCHPEQOPQNGSVTCFGEADQCVACAHYKDPFCVACRPS 602
 DB 538 KCNILEGEPREVENSECICQHEPLCPQANNTICTGRGPDNCICQAHYIDGHCVKTCFA 597
 QY 603 GYKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKCPAEQASPLTSTVSAVVGIL 662
 DB 598 GIMGENNTL-VWKYADANNVCHLCHANCTYGCAGPLOGCEVMPSPKIPSTATGIVGGL 656
 QY 663 LVVVLGWVFGI-LIKRQOKIRKYTMRLLQTELVEPLTPSGAMPNOAMRILKETEURL 721
 DB 657 LFIWV-VALGIGLFMRRRHIVKRTLRLQLRELVEPLTPSGEAPNQAHLRIKETEFPK 715
 QY 722 KVKVLGSGAFGTYYGIMIPDGENVKIPVAIKVLRNTSPKANKELIDRAYVMAGVGSY 781
 DB 716 KIKVLGSGAFGTYYKGLMIPGEKVKIPVAIKELREATSPKANKELIDRAYVMASVDNPH 775
 QY 782 VSRLLIGICLTSTVQLVTOLMPYGCLLDHVRENKRLGSDLLNCWCQIAKMSYLEDVRL 841
 DB 776 VCELLIGICLTSTVQLITQLMPYGCLLDYVREKDNIGSQYLLNVCVQIAKGNVLEDRRL 835
 QY 842 VHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMKMALESILRRRFFTH 901
 DB 836 VHRDLAARNVLKTPQHKVITDFGLAKLLGABEKEYHAEGGKVPKMKMALESILHRIYTH 895
 QY 902 QSDVWSYGYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVWKCM 961
 DB 896 QSDVWSYGYVWELMTFGSKPYDGIIPASDISILEKGERLPQPPICTIDVYIMVWKCM 955
 QY 962 DSECPRPRELVSFESRMARDPQRFVIO-NEDLGPASPLDSTFYRSLLEDGMDGLVDA 1020
 DB 956 DADSPKRELLLEFSQWARDQRYLVIQGDERMHPSTDSNFYALMEDMEDVDA 1015
 QY 1021 EBYLVPOQGFPCPDPAAGAGGMVHRRSSSTRSGGDI/TLGLPSEESAPSPAPSG 1080
 DB 1016 DEYLVPOQGF-----NSPST-----SRTPLLSLS 1041
 QY 1081 AGSDVFDGDLGMAKGLQSLTHDPSPLORYSEDPVLPSET--DGIVAPLTCSPQE 1138
 DB 1042 ATSN----NSTVACINRNGSCRVKEDAFLORYSSDPTGAVTEDNIDDAFL-----PVPE 1091
 QY 1139 YVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKGVVVKDVAFGGAVERNPEYL- 1197
 DB 1092 YVNPQ-SVPKRPAGSVQNPVYHNQPLHP-----APGDLHYQN--PHSNAVGNPEYLN 1140
 QY 1198 TPQGAAPQPHPPPAFSPAFDNLVYWDQ-----DP-----PERGAPSTFKGTP 1241
 DB 1141 TQA-----PTCLSSGFNSPALWIKGSHOMSLDNDPDYQDFFPKETKNGIFKG-P 1190
 QY 1242 TAENPEYLGLDVP 1254
 DB 1191 TAENAEYLRVAPP 1203

A;Accession: A27720
A;Molecule type: mRNA
A;Residues: 1-1223 <LAX>
R;Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A;Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: A00643
A;Molecule type: mRNA
A;Residues: 585-1223 <NIL>
A;Cross-references: GB:M10066
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F;31-654/Domain: extracellular #status predicted <EXT>
F;81-307/Domain: EGF receptor extracellular domain repeat <EE1>
F;397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F;655-677/Domain: transmembrane #status predicted <TM>
F;678-1223/Domain: intracellular #status predicted <INT>
F;719-984/Domain: protein kinase homology <KIN>
F;727-735/Region: protein kinase ATP-binding motif
F;136,202,360,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #s
F;192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F;687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;754/Active site: Lys #status predicted
F;1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat
Query Match 44.8%; Score 3052.5; DB 1; Length 1223;
Best Local Similarity 48.3%; Pred. No. 9.1e-118;
Matches 627; Conservative 171; Mismatches 353; Indels 147; Gaps 27;
QY 8 RGLLLALLPPGAA-----STQVCTGTMKRLPASPETHLDMLRHLHYGCGVQVGNLE 61
DB 13 RGAALVLLVLLGVALCSAVEEKVCQTNNKLTQLGHVEDHFTSLQRMYNCEVLSNLE 72
QY 62 LTVLPNASLSLQDIQEVGYVLIHNVQVPLQRLRIVRGTQLFEDNYALAVLDNGD 121
DB 73 ITVEHNRLDTLTKTQEVAGYVLIHNVQVPLQRLRIVRGTQLFEDNYALAVLSNYH 132
QY 122 PLNNTPTVTCASPGGLRLQLRSLRLILKGGVLIORNPOLCYDTLLWKDIFHKNQAL 181
DB 133 -MNKTQ-----GLRELPMKLSLNGVVKISNNPKLNDMTVLNDITDTSK-PL 182
QY 182 TLID-TNRSRACHPCSPMKSGRCWGESSEDCQSLRTRTVACGCA-RCKGPLPTDCCHQ 239
DB 183 TVLDFASNLSSCPKCHPNCTEDHCWGAGQCQTLTKVCAQCSCRCRGKVPSCCHNQ 242
QY 240 CAAGCTGPKHSCCLACHFNHSGICELHCPALVYNTDIFESMPNPEGRYTGASCVTAC 299
DB 243 CAAGCTGPKRESCLACRFRDDATCKDTPPLVLYNPTTYQMDVNPGEKYSFGATCRC 302
QY 300 PYNLYSTDVGSCTLVCPHLNQEVYARDGTORCKSKPCARVCYGLGMVYKANSKFIGI 359
DB 303 PHNVVTDHSGVSRSCNTTYEV-FENGVRKCKCDGLSKVCNGIGIGELKGILS-INA 360
QY 360 TELE-FAGCKKIFGSLAFIPESPDGPASNTAPQYIKANSKFTGITELTYGLYISAWPS 418
DB 361 TNIDSFNCTKINGDVSILPVAFLGDAFTKPLDPKLDVFRVKEISGFLLIQWPDN 420
QY 419 LPDLSVFQNLQVIRGILHNGAYSILTQGLISWGLRLSLRELGSGLAIHHTHLCFVH 478
DB 421 ATDLYAFENLEIRGTQKHQGYSLAVNLKIQSLGLSLKLSKESDGDIAIMKNKNLCYAD 480
QY 479 TVPWDDLFNPHQALLHTANRPEDECVGSLACHOLCARGHCWGPGPTOCVNCQSLRQ 538
DB 481 TNWRSLSFATQSKTKLIQNRKNKNDCTADRHVCDPLCSDVGCWGGPGFFHCFSRFRSQ 540
QY 539 ECVECRVLQGLPREYVYNARHCLPCHPECPQNG----SVTCFGEADQCACAHYKDPFF 595

DB 541 ECVKQCNILQGEPRFERDSKCLCHSECLVQNSTAYNTTSCGPGDHCMKAHFTDGP 600
QY 596 CVARCPGKVPDLSPYMPYTKWFPDEBACOPCPINCHSCVDLDDKGCPAEQRASPLTSIV 655
DB 601 CVKACPAGVGENDTL-VKMYADANAVCOLCHPNCTRGCKGPGLEGCP---NGSKTPSTA 656
QY 656 SAVV-GILYVVVLGVVFGILIKRQOKIRKYTRMRLLQETELVEPLTPSGAMPNQOMRI 714
DB 657 AGVVGGLLCLVVVGLGIGLYLRRR-HIVKRTLRLLQERLVEPLTPSGEAPNQOHLRI 715
QY 715 LKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTSPKANKELIDAYVM 774
DB 716 LKETELFKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTSPKANKELIDAYVM 775
QY 775 AGVGSPPYVSRLLGICLTSTVQLVLTQMPYGCLLDHVRENRRGLSGDILNWCWQIAKMS 834
DB 776 ASVDNPHVCRLLGICLTSTVQLVLTQMPYGCLLDHVRENRRGLSGDILNWCWQIAKMN 835
QY 835 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALES 894
DB 836 YLEERRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALES 895
QY 895 LRRRTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVMI 954
DB 896 LHRITHTQSDVWSYGVTVWELMTFGSKPYDGIIPAREIPDLLEKGERLPPOPICTIDVMI 955
QY 955 MYCKWMIDSECPREPRFELVSEFMRARQPRFVVIQ-NEDLGPASPLDSTFVRSLLDD 1013
DB 956 MYCKWMIDSECPREPRFELVSEFMRARQPRFVVIQ-NEDLGPASPLDSTFVRSLLDD 1015
QY 1014 MGDVDADEYLVPQOQFCPPDPAAGGVMVHRHRSSTSRGGDILTLGLBPESEAPRS 1073
DB 1016 MEDIVDADEYLVPQOQFCPPDPAAGGVMVHRHRSSTSRGGDILTLGLBPESEAPRS 1073
QY 1074 PL-----APSEGAGSDVFDGLGMAAGKGLQSLPHDPSPLQRYSEDPTVPLPSET--DG 1126
DB 1042 PLLSLSATSNNNSATNCID-----RNGQGHVREDSFVQRYSSDPTGNFLESIDG 1093
QY 1127 VYAPLTCSPQRPVNOVPOVPOPPSPREGPLPAARAGATLERAKTSLSPKGVVKKVDF- 1185
DB 1094 FL-----PAPVYVQ--LMPKKPS-----TAMVQVYNNISL 1124
QY 1186 -----AFGGAIVENPEYLTQGGAAPOHPHPAPSPAFDNLVYDQ----- 1225
DB 1125 TAISKLPWDERVQNSHSTAVDNPEYL-----NTNQSPLAKTVFESSPYIQSGNHOI 1176
QY 1226 --DPPE-----RGAPPSTFKGTPTAENPEYLGIDVP 1254
DB 1177 NLNDNPYQOQFLPNETKPNGLLKVPAAENPEYLRVAAP 1214
RESULT 7
A47253
epidermal growth factor receptor, HER4 - human
C:Species: Homo sapiens (man)
C>Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C:Accession: A47253
R:Plowman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; N
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A:Title: Ligand-specific activation of HER4/pl80erbB4, a fourth member of the epidermal
A:Reference number: A47253; MUID:93189574; PMID:8383326
A:Accession: A47253
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1308 <PLO>
A:Cross-references: GB:L07869; NID:g337359; PIDN:AAB59446.1; PID:g337360
A>Note: sequence extracted from NCBI backbone (NCBI:126842)
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor
F;716-981/Domain: protein kinase homology <KIN>
F;724-732/Region: protein kinase ATP-binding motif

Db 123 YQK-NPSSP--DYVQVGLKQLQLSNLTELISGKVSHNPLLCNVETINWWDIVDKTNP 179
QY 180 ALTLIDNRSRACHPCSPMCKGRWCWGESSEDCQSLTRTVACAGC-ARCKGPLEPTDCHE 238
Db 180 TNNLIIPAFERQCKQCDHGCVNGSWAPGPGHGCQKFTKLLCABQCNRRCRGPKPIDCCNE 239
QY 239 QCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA 298
Db 240 HCAGGCTGPRATDCLACRDNFDGTCKTCCPPKIYDIVSHQVVDNPNKTYTFGACVAKVE 299
QY 299 CPXNYLSTDVGSCTLVCPHLNQHVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKFTG 358
Db 300 CPXNYVVTB-GACVRSAGMLEVD-ENGKRSCKPCDGVCPKVCDDIGIGSL-SNTIAVN 356
QY 359 ITL-EPAGCKITFGLAFIPESFDGP---ASNTAPQYIKANSKFTIGITGLYLYISA 414
Db 357 STNIRSFNCKINGDIIILNRNSFEGDPHYKIGTMDPEHL---WNLTTVKEITGYLVIW 413
QY 415 WPSLPLDLPVQNLQVIRGRILHNGAYS-LTLOGLGISWLGRLSRLSELGSLALIHNTH 473
Db 414 WPNMTSLSVFQNLRIIRGRTTFSRGFSFVVQVVRHLQWLGRLSKLSEVAGNVLKNTLQ 473
QY 474 LCFVHTVPMDLPFRNPHQALLHTANRPEDECVGEGLAGHOLCARGHCWGPPTQCVNCSQ 533
Db 474 LRYANTINWRLEFRSEDQIEYDART-----ENQTCNNEGSEDCGW-PGPTMVCVSLH 525
QY 534 FLRGQCEBCEVRLOGLPREYVNAHCLPCHPECPQNGSVTFCGPEADQCACAHYKDP 593
Db 526 VDRGRCVASCNLLQGEPREAQVDRGVCVQCHQCLVQTSLTCTCYGPGPANCSSAHFQDG 585
QY 594 PFCVAPCGVGRKPLSYMIWPKPDEEGACQPCPNCTHSCVDLDDKGPCPAEQRASPLTS 653
Db 586 PQCIPCPHGILGDGDTL-IWKYADRWGQCQPCQNCQTCQCSGPGSLSGCRGD-IVSHSL 643
QY 654 IYSAVVGILLVVLGVVFGILIKRRQKIRKYTMRELLOETELVEPLTSGMENAQWR 713
Db 644 AVGLVSLGITVIVALLIVLLVLRRIK-RKRTIRCLQELKELVETLPSGQAPNAR 702
QY 714 ILKETLRKVKVLGSGAFGVYKGIWIPDENVKIPVAIKVRENTSPKANKEILDEAYV 773
Db 703 ILKETEFKDRVLGSGAFGVYKGLWNPDENIRIPVAIKVLEATSPKVQNEVLDEAYV 762
QY 774 MAGVGPYVSRLLIGICLTSTVQLVTQMPYGCGLLDHVRNRRGLSGQDLNWCWQIAKM 833
Db 763 MASVDHPHVCRLGICLTSAVQLVTQMPYGCGLLDYVROHERICQGWLLNWCWQIAKM 822
QY 834 SYLEDVRLVHRDLAARNLVKSPNHVKITDGLARLLDDEYHADGKVPKKNWALS 893
Db 823 NYLEERHLVHRDLAARNVLLKNPNHVKITDFGLSKULLTADKEKYQADGKVPKKNWALS 882
QY 894 ILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYM 953
Db 883 ILQWYTHQSDVMSYGVTVWELMTFGSKPYDGIPIAKEIASVLENGERLPQPPICTIEVM 942
QY 954 IMVKWMIDSECRPRELVSFSPRMWARDPQRPVVTIONEDLGPASPLDSTFYRSLLEDD 1013
Db 943 IILKWMIDPSSRPRELVSFSPRMWARDPSRVLVIGQ---NLPSLSDRRLFSRLSSD 999
QY 1014 MGLVDABEVLVPOQGFCDPAPAGAGVHHHRSSSTRSGGDDTLTGLESEEAAPS 1073
Db 1000 --DVVDADVELLPYKRI-----NRQS-----E 1020
QY 1074 PLAPSEGAGSDVPDGLGMAAKGLQSLPHTDPSLQRYSEDPV-PLPSETDGVVAPLT 1132
Db 1021 PCIPPTGH-----PVRENSITRLNISDPTQNALSKDLGH----- 1055
QY 1133 CSPQPYVNPQDVRPQ-----PSPRE-----GFLP-AARPAGATLERAKTLSPGKNG 1179
Db 1056 -----EYVNPQSETSSRLSDIYNPNYEDLTGCGVPVSLSSQEAETNFSRPEYINTNQS 1110
QY 1180 VVKDVFAGAVENPEYLPFOGGAAPQPHPPAFSPAFNLYWDQDPPERGAPPTFKG 1239
Db 1111 L-----PLVSSGSMDDPDP---QAG-----YQAAF-----LPQTGALTGNGMF 1145

QY 1240 TPTAENPEYL3 1250
Db 1146 LPAEENLEYLG 1156

RESULT 9

A36223

Kinase-related transforming protein (erbB3) (BC 2.7.1.-) precursor - human
Species: Homo sapiens (man)

C>Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000

C.Accession: A36223; 159164

R:Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.

Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989

A:Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal

A:Reference number: A36223; MUID:90083234; PMID:2687875

A.Accession: A36223

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1342 <KRA>

A:Cross-references: GB:M29366

R:Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, G.

Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990

A:Title: Molecular cloning and expression of another epidermal growth factor receptor-

A:Reference number: 159164; MUID:90311312; PMID:2164210

A.Accession: 159164

A:Status: preliminary; translated from GB/EMBL/DBDJB

A:Molecule type: mRNA

A:Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>

A:Cross-references: GB:M34309; NID:g183990; PIDN:AAA35979.1; PID:g306841

C:Genetics:

A:Gene: GDB:ERBB3; HER3

A:Cross-references: GDB:119880; OMIM:190151

A:Map position: 12q13-12q13

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homol

C:Keywords: ATP; phosphotransferase

F:707-972/Domain: protein kinase homology <KIN>

F:715-723/Region: protein kinase ATP-binding motif

Query Match

Best Local Similarity 34.7%; Score 2366.5; DB 2; Length 1342;

Matches 524; Conservative 197; Mismatches 458; Indels 135; Gaps 34;

QY 10 GLLALLPPGAA--STQVCTGTMKRLPASPETHLMLRHLHYQGVVQGNLELYLPT 67

Db 11 GLLFSLARGSEVGNSSQAVCPGTLNGLSVTDAENQVQTLKLYERCEVVMGNLEIVLTGH 70

QY 68 NASLSFLQDIQEVQYVLIHNOVROVFLORLIRVIRGTQOLFEDNYALAVLDNDPLNNTT 127

Db 71 NADLSFLQWIREVTGYVLVAMNEFSTLPLNLRVVVRGTQVYDGKFAIFVM----LNynt 125

QY 128 PVTGASPGGLRELQLRSLEILLKGGVLIQRPOLCVQDTILWKDIFHKNNQLALTIDTN 187

Db 126 ----NSSHALRQRLTQTEILLSGGVYTEKDKLCHMDIIDWRDIVRDRD---AEIVKD 178

QY 188 RSRACHPCSPMCKGRWCWGESSEDCQSLTRTVACAGC-ARCKGPLEPTDCHEQCAAGCTG 246

Db 179 NGRSCPPCHEVCKG-RWGPGEEDCQTLTKITCAPQCNGHCFGPNPQCCHDECAGGCSG 237

QY 247 PKHSCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAAPNYLST 306

Db 238 PQDTCFACRHFNDSGACVPRCPQPLVYVKNLTFQLEPNPHTKYQYGGVGVASCNHFV-V 296

QY 307 DVGSCCTLVCPHLNQHVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKFF--IGITELB- 363

Db 297 DQTSVCRACPPDNMEVD-KNGLKMCPCGGLCPKACEGTG-----SGSRFQTVDSNIDG 350

QY 364 FAGCKKIFGSLAFPESEFDGPPASNTAPQYIKANSKFTIGITELTGYLYISAWPDSLPLDS 423

Db 351 FVNCYTKILGNLDFLTGLNGDPWHKI PALDPEKLVNFRVREITCYLNIQSWPPHMFPS 410

QY 424 VFQNIQVIRGRILHNGAYS-LTLOGLGISWLGRLSRLSELGSLALIHNTHLCHFVHTVPW 482

```

Db 411 VFSNLTTIGRSLYNRGFSLLIMKNLNVTSGLFRSLKEISAGRIYISANRQLCYHSLNW 470
Qy 483 DQLFRPHQALLHTA-NRPEDECVGGLACHOLCARGHCWGPGTQCVNCSQFLRQECV 541
Db 471 TKVLRGTEERLDIKHNRPRDCVAGKVCDDPLCSSGGCGWPGPGQCLSCRNYSRGVCV 530
Qy 542 EECRLVQLGLPREYVYVNAHCLHPCEQOPQNGSVTCFQPEADQCVACAHYKDDPPFCVARCP 601
Db 531 THCNFLNGEPREFAHEACFSCHPEQOPMEGTATCNGSGSDTCAQCAHFRDGHCVSSCP 590
Qy 602 SGVKPDLSPYMPIWKFDPDEBAGQCPINCHTHSCVDLDKGCAPAEQRA----SPLTIVSA 657
Db 591 HGVLG--AKGPIYKYEDVQNECPCHENCCTQCGKGPPELQCLGQTLVLIGKTHLTALTY 648
Qy 658 VVGLLVVVLGVVFGILIKRQOKIR-KYTMRELLQETELVEPLTSGAMPNQAQWILK 716
Db 649 IAG--LVVIFMGLGTFVYWRGRRIQNKRAMRYLRGEGSIEPLDPS-EKANVKVLARIFK 705
Qy 717 ETELRLKVLGSGAFVTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDYAVMAG 776
Db 706 ETELRLKVLGSGVFGTVHKVMIPEGESIKIPVCIKVIEDKSGROSFOAVTDHMLAIGS 765
Qy 777 VGSPPYVRLIGICLTSTVQLVLTQMPYVGLLDHVRNENRGLSGQDLNLMWCMQIAKMSYL 836
Db 766 LDHAHIVRLGLCPGSSQLVTVQYPLGLSLLDHVRQHRGALGPQLLLNMGVQIAKMWYL 825
Qy 837 EDVRLVHRLAARNVLKSPNHVKITDPLGLARLLDIDETEHADGGKVPILKNWALESLR 896
Db 826 EEHGMVHRNLAARNVLKSPQVQVADFGVADLLPDDKQLLYSEAKPTIKNWALESIHF 885
Qy 897 RRTTHOSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTTIDVYMIW 956
Db 886 GKTHQSDVMSYGVTVWELMTFGAEYAGRLAEVDPDLLEKGERLAQPOICTTIDVYMWV 945
Qy 957 KCMWIDSECRPRELVSEFSRMARDPQREVVTIONEDLGA----SPLDSTFYHSLLEDD 1013
Db 946 KCMWIDENIRPTEKELANEFTRMARDPPRYLVIKRES-GPIAPGPEPHGLTNKLEEEV 1004
Qy 1014 MGLVDAAEEVLVPOQFFCPDPAPAGGMVYHHRSSSTRSGGDLTLGLEP-SESEAR 1072
Db 1005 LEPELDLDLLEAED-----NLATTGLSALSLPVTGLNLRPRGSQ 1045
Qy 1073 SPLAPSEGAGSDVFDGLGMAKGLQLPHTD-PSPLQRYSEDPTVPLP-----SETD 1125
Db 1046 SLSPSSGY-MPMNQNLGSECOESAVSGSERCPRVSLH-----PMRGLGASESE 1098
Qy 1126 GYVA-----PLTCSPOE----YVNPQDVVRPQPPSPREGP----- 1156
Db 1099 GHVTGEAELOEKVSMCRSRSRSPRPRGDSAYHSQRHSLLTPVTPLSPPGLEBEDVNG 1158
Qy 1157 --LPAARPAGATLERAKTLSP-GKNGV-----KDVFAFGGAVENPEYLTPOGGAPOP 1207
Db 1159 YNMPDTHLKTSPSREGTUSVGLSSVLGTEEBED-----EEYENRRRRHSP-P 1209
Qy 1208 HPPAPSPADNLYYWD-----ODPPERGAPPSTFTKGTPTAENPEYL 1249
Db 1210 HPPRPSLEELGYEYMDVGDLSASLGSTQSCPLHPVIMPATAGTTPDEDEYEM 1263

```

RESULT 10
JC4387

epidermal growth factor receptor; homolog precursor - rat

N;Alternate names: ErbB3 protein; HER3 protein

C;Species: Rattus norvegicus (Norway rat)

C;Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998

C;Accession: JC4387

R;Hellyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.

Gene 165, 279-284, 1995

A;Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.

A;Reference number: JC4387; MUID:96096535; PMID:8522190

A;Accession: JC4387

A;Molecule type: mRNA

A;Residues: 1-1339 <HEL>

A;Cross-references: GB:U29339; MID:g915389; PID:g915390
A;Experimental source: liver
A;Note: The authors translated the codon AAC for residue 369 as Thr and CTT for residue
C;Comment: this protein is a functional heregulin receptor that transduces signals to ti
C;Genetics:
A;Gene: ErbB3
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homol
C;Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1339/Product: epidermal growth factor homolog #status predicted <TM>
F;640-659/Domain: transmembrane #status predicted <TM>
F;705-970/Domain: protein kinase homology <KIN>
F;713-721/Region: protein kinase ATP-binding motif
F;939, 1051, 1156, 1194, 1196, 1219, 1257, 1259, 1273, 1286, 1325/Binding site: phosphate (Tyr)

Query Match 33.5%; Score 2280.5; DB 2; Length 1339;

Best Local Similarity 40.1%; Pred. No. 3.6e-86;

Matches 515; Conservative 173; Mismatches 438; Indels 159; Gaps 35;

```

Qy 3 LAALCRWGLLLALLPPGAA---STOVCTGTDMKRLRPASPETHLDMLRHLHYQGCQVQGN 59
Db 7 LQVLC----FLSLARGSEMNSQAVCPGTNLGLSVTGADNQYQTLYKLYEKCEVVMGN 62
Qy 60 LELTYLPTNASISFLQDIQEVQGYVLIHQNVRQVPLQRLRIVRGTLQFEDNYALAVLN 119
Db 63 LEIVLTGHMADLSFLQWIREVTAYVLVAMNFSVLPPLNLRVVRGTQVYDGKFAIFVM-- 120
Qy 120 GDLNNTPTVTCASPGGLRELQLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKQNQL 179
Db 121 ---LNYNT---NSSHALRQLKFTQLTEILSGGVVIEKNDKLCMDTIDMRDIVRVR--- 170
Qy 180 ALTLDTNRSRACHPCSPMCKSGWGESSEDCQSLTRTVACAGG-ARCKGLPLPTDCCHE 238
Db 171 GAIVVYKANGANCPPCHEVCKG-RCWGPDPDCCQLTKTICAPQNGRCFGPNPNOCCHD 229
Qy 239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFFSMPNPEGRYTFGASCVTA 298
Db 230 ECAGSGSGQDPTDFACRRFNDSGACVPRCPPEPLVYNKLTFLQLEPNPHTKYQVGGVCVAS 289
Qy 299 CPYNYLSTDVGCTLVCPLNQEVTAEDGTQRCCKSRPCARVCYL--GMQYIKANSKF 356
Db 290 CPNFFV-VDTQFCVRACPPDKMEVD-KHGLKNCEPGGLCPKACEGTGSGSRVQTVDDSN 347
Qy 357 IGI TELEFAGCKKIFGSLAFLPESFDGDPASNTAQYIKANSKFTIGITELTGVLYISAWP 416
Db 348 ID----GFVNCYKILGNLDFLTGLNVDPWHKIPALDPEKLVNFRVTRITGLNIQSWP 403
Qy 417 DSLPDLVSFQNLQVIRGRILHNGAYS-LTLOGLGISWGLRSLRSLRGLSGLALIHNTILC 475
Db 404 PHMNFVSFNLTTIGRSLYNRGFSLLIMKNLNVTSGLFRSLKEISAGRVVISANQQLC 463
Qy 476 FVHTVPMDQLFRPHQALLHTA-NRPEDECVGGLACHOLCARGHCWGPGTQCVNCSQF 534
Db 464 YHHSNLNWTLLRGPSEERLDIKYDRPLGCELAEGKVCDDPLCSSGGCGWPGAPQCLSCRY 523
Qy 535 LRQSCVEECRVQLPREYVYVNAHCLPCHPSCQPNQSVTCFQPEADQCVACAHYKDDP 594
Db 524 SREGVCVTHCNFLOGEPRFVHEAQCFCHPCLPMEGTSTYNGSGSDACACAHPRDP 583
Qy 595 FCVARCPSGVKPDLSPYMPIWKFDPDEBAGQCPINCHTSC--VDLDDKGCAPAEQASPLT 652
Db 584 HCVNSCPHGILG--AKGPIYKYPDQANECPCHENCCTCGNGPELQCLGQAEVLMSKPH 641
Qy 653 STVSAVVGLLLVVLGVVFGIILIKRQOKIR-KYTMRELLQETELVEPLTSGAMPNQAQ 711
Db 642 LVIAVTVG--LAVILMILGSGFLYWRGRRIQNKRAMRYLRGEGSIEPLDPS-EKANVKL 698
Qy 712 MELKETELRKVKVLGSGAFVTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDYEA 771
Db 699 ARIFKETELURKVLGSGVFGTVHKGIWIPGESIKIPVCIKVIEDKSGROSFOAVTDM 758
Qy 772 YVMAGVGSPPVSRLLIGLICLTSTVQLVLTQMPYVGLLDHVRNENRGLSGQDLNLMWCMQIAK 831

```


Matches 345; Conservative 69; Mismatches 122; Indels 92; Gaps 15;

```
QY 578 GPRADQVCAHYKDPFCVAPCPGVKPDLSYMPIWKFPDEBACQPCPINCINCHSCVDL 637
Db 1 GP--DHCMKCAHFDGPHCVKACPAAGVLGENDTL-VMKYADANAVCOLCHPNCTRGCKGP 57
QY 638 DDGCGCAEQRASPLTSTVSAAV-GILLVVVLGVVFGILLIKRRQOKIRKYTMRLLOETEL 696
Db 58 GLEGGCP---NGSKTPSIAAGVVGGLCLVVGGLGIGLYLRR-HIVKRTLRLLQEREL 113
QY 697 VEPLTPSGAMPNOAQRILKTELRLKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRL 756
Db 114 VEPLTSGEAPNOAHLRIKTEFKVKVLGFGAFGVYKGLWIPGEKVTIPVAIKELR 173
QY 757 ENTSPKANKEILDEAYMAGVSGPYVSRLLIGICLTSTVQLVTLQMPYGCLLDHVRENRR 816
Db 174 EATSPKANKEILDEAYMASVDNPHVCRLLIGICLTSTVQLITQMPYGCLLDIYREHKON 233
QY 817 LGSODLLNCMOIAKMSYLEVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 876
Db 234 IGSQYLLNCVQIAKGMNLEERHVLHRDLAARNVLKTPQDVKITDFGLAKQLGADEKE 293
QY 877 YHADGGKVPPIKMALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLE 936
Db 294 YHAEKGKVPPIKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISVLE 353
QY 937 KGERLPQPTCTIDVYIMVKMWIDSECRPRFRELVSFMRMARDPQRFVVIQ-NEDLGP 995
Db 354 KGERLPQPTCTIDVYIMVKMWSDADSRPKFRELIAEFKSMARDPPRYLVIQGDERMH 413
QY 996 PASPLDSTFVRSLEDDMGDLVDAEYLVPOQGFPCPDAPGAGGMVHRRHSSSTRSG 1055
Db 414 LPSPTDSKFTVRLMEEDMEDIVDAEYLVPHQGF-----NSPST---454
QY 1056 GGDLTGLPSEBEAPRSL-----APSEGAGSDVFDGDLGMGAAGLQSLPHTDPSPLQ 1110
Db 455 -----SRTPLLSLSATSNNSATNCIDRNG-----H-----481
QY 1111 RYSEDTVPLPSTGDYVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAGAT-LER 1169
Db 482 -----PVREDGFL-----PAPEYVNQ--LMPKPKSTAMVQNIYVYISLTAISK 523
QY 1170 AKTLPKGVKGVKDVFAFGAVENPEYL 1197
Db 524 LPMDSRYQ-----SHSTAVDNPEYL 544
```

RESULT 15

B44776

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain ES4)

C;Species: avian erythroblastosis virus

C;Date: 28-Apr-1993 #sequence_revision 28-Apr-1993 #text_change 04-Feb-2000

C;Accession: B44776

R;Bruskin, A.; Jackson, J.; Bishop, J.M.; McCarley, D.J.; Schatzman, R.C.

Oncogene 5, 15-24, 1990

A;Title: Six amino acids from the retroviral gene gag greatly enhance the transforming p

A;Reference number: A44776; PMID:90206603; PMID:1969616

A;Accession: B44776

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-540 <BRU>

A;Cross-references: GB:X52211

C;Genetics:

A;Gene: erbB

C;Superfamily: epidermal growth factor receptor; protein kinase homology

C;Keywords: ATP; phosphotransferase; tyrosine-specific protein kinase

F;130-395/Domain: protein kinase homology <KIN>

F;138-146/Region: protein kinase ATP-binding motif

Query Match

Best Local Similarity 54.9%; Pred. No. 1.1e-59; Length 540;

Matches 340; Conservative 69; Mismatches 120; Indels 90; Gaps 14;

Search completed: July 22, 2003, 09:09:39

Job time : 30.9062 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 37.9774 Seconds
(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-250-264-12
Perfect score: 6801
Sequence: 1 MELALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6720	98.8	1255	21	AAV92620
2	6720	98.8	1255	22	AAE12130
3	6720	98.8	1255	22	AAE60167
4	6720	98.8	1255	23	AAU74545
5	6714	98.7	1255	17	AAW01111
6	6714	98.7	1255	20	AAW92406
7	6714	98.7	1255	21	AAE21198
8	6714	98.7	1255	21	AAV84780
9	6714	98.7	1255	22	AAE85458
10	6714	98.7	1255	22	AAE88267

11	6714	98.7	1255	23	AAE24067	Human Her-2 protei
12	6714	98.7	1255	23	AAE20479	Human Her-2/neu pr
13	6714	98.7	1255	23	AAE51143	Human Her-2/neu on
14	6714	98.7	1255	23	AAU77114	Human Her-2/neu po
15	6671	98.1	1433	14	AAE39568	Sequence of c-erbB
16	6550	96.3	1223	23	AAU98923	Human breast cance
17	6397	94.1	1200	21	AAE21208	Human HER-2/neu pr
18	5915.5	87.0	1256	21	AAE21199	Rat HER-2/neu prot
19	5915.5	87.0	1256	23	AAE51144	Mouse Her-2/neu on
20	5889.5	86.6	1256	21	AAE21206	Mouse Her-2/neu pr
21	5889.5	86.6	1256	22	AAE62860	Amino acid sequenc
22	5889.5	86.6	1256	23	AAE51151	Mouse Her-2/neu on
23	4800	70.6	919	21	AAE21203	Human HER-2/neu fu
24	4800	70.6	919	23	AAE51148	Her-2/neu extracel
25	4039.5	59.4	920	23	AAE51152	Mouse Her-2/neu ex
26	4039.5	59.4	926	23	AAE51153	Mouse Her-2/neu ex
27	3684	54.2	712	21	AAE21204	Human HER-2/neu fu
28	3684	54.2	712	23	AAE51149	Her-2/neu extracel
29	3536	52.0	782	18	AAW19764	Her2-GM-CSF immuno
30	3536	52.0	653	21	AAE21200	Extracellular HER-
31	3536	52.0	653	23	AAE51145	Human Her-2/neu on
32	3498	51.4	645	22	AAE60408	Human ErbB2 oncopr
33	3498	51.4	645	22	AAE61593	Human ErbB2 extrac
34	3433	50.5	951	21	AAE44993	DC8ecFv-erbB2EC fu
35	3330	49.0	624	11	AAE08222	Extracellular port
36	3121	45.9	1210	21	AAE19259	Amino acid sequenc
37	3121	45.9	1210	21	AAE50616	Human EGF receptor
38	3121	45.9	1210	23	AAE23019	Human Her-1 protei
39	3121	45.9	1210	23	AAE50768	Human epidermal gr
40	3119	45.9	1210	22	AAE68420	Amino acid sequenc
41	3084	45.3	583	23	AAE20483	Human protein for
42	3084	45.3	587	23	AAE20481	Human protein for
43	3083	45.3	589	23	AAE20484	Human protein for
44	3083	45.3	600	23	AAE20482	Human protein for
45	3080	45.3	1210	23	AAE51768	Human epidermal gr

ALIGNMENTS

RESULT 1	
AAV92620	AAV92620 standard; Protein; 1255 AA.
ID	AAV92620
XX	AAV92620; (first entry)
AC	10-AUG-2000
DT	DT
XX	Human heregulin 2 (Her2).
DE	Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
XX	self-protein; cancer; breast cancer; prostate cancer;
KW	cell-associated peptide antigen; foreign epitope.
KW	
OS	Homo sapiens.
XX	
PH	Key
FT	Domain
FT	Location/Qualifiers
FT	1..173
FT	/label= N-terminal
FT	/note= "mature polypeptide"
FT	5..25
FT	/label= insertion region
FT	/note= "suitable for foreign epitope insertion"
FT	59..73
FT	/label= insertion region
FT	/note= "suitable for foreign epitope insertion"
FT	103..117
FT	/label= insertion region
FT	/note= "suitable for foreign epitope insertion"
FT	149..163
FT	/label= insertion region
FT	/note= "suitable for foreign epitope insertion"
FT	174..323
FT	Domain

FT Region /label= Cysteine_rich_domain
 FT 210..224
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 250..264
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 324..483
 FT /label= Ligand_binding_domain
 FT 325..339
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 369..383
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 465..479
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 484..623
 FT /label= Cysteine_rich_domain
 FT 579..593
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 624..654
 FT /label= Transmembrane_domain
 FT 632..652
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 653..667
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 655..1010
 FT /label= Tyrosine_kinase_domain
 FT 661..675
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 695..709
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 710..730
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 1011..1235
 FT /label= C-terminal_domain
 XX WO200020027-A2.
 XX 13-APR-2000. 99WO-DK00525.
 XX 05-OCT-1999;
 XX 05-OCT-1998; 98DK-0001261.
 XX 20-OCT-1998; 98US-0105011.
 XX (MEBI-) M & E BIOTECH AS.
 XX Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
 XX Gautam A, Birk P, Karlsson G;
 XX WPI; 2000-349917/30.
 XX N-PSDB; AAA09455.
 XX Inducing immune responses to weakly immunogenic, tumor associated
 XX peptide antigens for the treatment of breast and prostate cancer
 XX
 XX Claim 62; Page 193-198; 220pp; English.
 XX This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
 XX Her2 can be used in the claimed method as an autovaccine to induce a CTL
 XX response. Subdominant CTL epitopes, antibody binding regions and
 XX cysteine residues involved in disulfide bonds are preserved in the
 XX immunogenized forms. Regions suitable for the insertion of foreign T
 XX helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
 CC cell-associated peptide antigens (PA) such as those associated with
 CC cancers (self-proteins). e.g. human prostate specific membrane antigen
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
 CC The method comprises effecting simultaneous presentation by antigen
 CC producing cells (APCs) of the animals immune system of: (1) at least 1
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
 CC B-cell group derived from the cell-associated PA; and (2) at least 1
 CC first T-helper cell group which is foreign to the animal. Analogues of
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
 CC part of all known and predicted CTL and B-cell epitopes of the respective
 CC PA and including at least one foreign T helper epitope are also claimed.
 CC The method is used to treat prostate, prostate/breast or breast cancer
 CC when the PA is human PSM, FGF8b and Her2, respectively.
 XX
 XX SQ Sequence 1255 AA;

Query Match 98.8%; Score 6720; DB 21; Length 1255;
 Best Local Similarity 99.0%; Pred. No. 0;
 Matches 1242; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASQTGCTDMKRLPASPETHLDMLRHLYQGCVVQGNL 60
 DB 1 MELAALCRWGLLLALLPPGAASQTGCTDMKRLPASPETHLDMLRHLYQGCVVQGNL 60
 QY 61 ELTYLPTNASLSFLQDIQEVQGVLLAHNQVROVPLQRLIRVRGTOLFEDNYALAVLDNG 120
 DB 61 ELTYLPTNASLSFLQDIQEVQGVLLAHNQVROVPLQRLIRVRGTOLFEDNYALAVLDNG 120
 QY 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILMKDIFHKNNQLA 180
 DB 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILMKDIFHKNNQLA 180
 QY 181 LTLIDNRSRACHPCSPMCKGSRGWESSEDCOSLTRTVCAGGCARCKGPLPTDCCHQEC 240
 DB 181 LTLIDNRSRACHPCSPMCKGSRGWESSEDCOSLTRTVCAGGCARCKGPLPTDCCHQEC 240
 QY 241 AGCTGPKHSDCLACLFHFNHSGICELHCPALVOYIKANSKFIGITELRITFFGASCVTACP 300
 DB 241 AGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMNPGRITFFGASCVTACP 300
 QY 301 YNYLSTDVGSCTLVCPHNLQEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
 DB 301 YNYLSTDVGSCTLVCPHNLQEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
 QY 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFETLEEITGVLYISAWPDSL 420
 DB 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFETLEEITGVLYISAWPDSL 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLIGISWGLRSLRELGLSLALIHNNTHLCFVHTV 480
 DB 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLIGISWGLRSLRELGLSLALIHNNTHLCFVHTV 480
 QY 481 PWDQLFRNPQALLHTANRPEDECVGEGLAHQLCARGHCWGPGTQCVCNCSQFLRGQEC 540
 DB 481 PWDQLFRNPQALLHTANRPEDECVGEGLAHQLCARGHCWGPGTQCVCNCSQFLRGQEC 540
 QY 541 VVECRVLOGLPREYVYNARHCLPCHPECQPNQSVTCFGEADOCVCAHYKDPFPCVARC 600
 DB 541 VVECRVLOGLPREYVYNARHCLPCHPECQPNQSVTCFGEADOCVCAHYKDPFPCVARC 600
 QY 601 PSGVKPDLSPYMPKPFDEEGACQPCPINCTHSCVDLDDKGCFAEORASPLTISVAVVG 660
 DB 601 PSGVKPDLSPYMPKPFDEEGACQPCPINCTHSCVDLDDKGCFAEORASPLTISVAVVG 660
 QY 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRLLOQETELVEPLTPSGAMPNQAQMRILKETEL 720
 DB 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRLLOQETELVEPLTPSGAMPNQAQMRILKETEL 720
 QY 721 RKVKVLGSAFGTVYKGIWIPGENVKIPVAIKVLRNTSPKANKELDEAYVMAGVGP 780
 DB 721 RKVKVLGSAFGTVYKGIWIPGENVKIPVAIKVLRNTSPKANKELDEAYVMAGVGP 780

QY 781 YVRLGICLTSTVQLVTLQMPYGCCLLDHVRNRRGLSGDILLNWCQIAKGSYLEYDVR 840
Db 781 YVRLGICLTSTVQLVTLQMPYGCCLLDHVRNRRGLSGDILLNWCQIAKGSYLEYDVR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEYHADGGKVPKIMWALESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEYHADGGKVPKIMWALESILRRRFT 900
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPCTIDVYIMVKCWM 960
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPCTIDVYIMVKCWM 960
QY 961 IDSECPRRRELIVSESRMARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
Db 961 IDSECPRRRELIVSESRMARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
QY 1021 EYLVVQQGFCDPAPGAGWVHRRSSSTRSGGDLTLGLPESEERAPRPLAPSEG 1080
Db 1021 EYLVVQQGFCDPAPGAGWVHRRSSSTRSGGDLTLGLPESEERAPRPLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAPGGAVENPEYLTQP 1200
Db 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAPGGAVENPEYLTQP 1200
QY 1201 GGAAPQHPPPAFSPAFDNLVYWDQPPPERGAPPSTFKGPTAENPEYLGLDVVPV 1255
Db 1201 GGAAPQHPPPAFSPAFDNLVYWDQPPPERGAPPSTFKGPTAENPEYLGLDVVPV 1255

RESULT 2
AAE12130
ID AAE12130 standard; Protein; 1255 AA.
XX AAE12130;
AC AAE12130;
DT 18-DEC-2001 (first entry)
XX Human tyrosine kinase-type receptor, HER-2.
DE
XX Therapeutic compound; major histocompatibility complex; vaccine;
KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
KW antigen presenting cell; human; tyrosine kinase-type receptor.
XX
OS Homo sapiens.
XX
Key Location/Qualifiers
FH Region 774..782
FT /note= "Antigenic epitope"
FT
XX WO200168677-A2
XX
XX 20-SEP-2001.
XX
XX 16-MAR-2001; 2001WO-US40328.
XX
XX 16-MAR-2000; 2000US-0527487.
PR
XX (GENZ) GENZYME CORP.
XX
XX Nicolette CA;
XX
XX WPI; 2001-616284/71.
DR N-PSDB; AAD19731.
XX
XX Novel synthetic therapeutic compound for inducing immune response and
PT for use in adoptive immunotherapy, has enhanced binding to major
PT histocompatibility molecules and enhanced immunoregulatory properties
PT

XX
PS
XX
CC The invention relates to synthetic therapeutic compounds (antigenic
CC peptides) with enhanced binding to major histocompatibility complex
CC (MHC) molecules and enhanced immunoregulatory properties relative
CC to their natural counterparts. Compounds of the invention are useful
CC for inducing an immune response in a subject and for use in adoptive
CC immunotherapy. They are useful as components of anti-cancer vaccines
CC and to expand immune effector cells that are specific for cancers
CC characterised by expression of the breast cancer antigen, HER-2.
CC Polynucleotides that encode peptides of the invention are useful as
CC hybridisation probes and as primers for the detection of genes of gene
CC transcripts that are expressed in antigen presenting cells (APCs), to
CC confirm transduction of polynucleotides into host cells. The present
CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
CC of the invention are designed based on the HER-2 antigenic peptide
CC (774-782).
XX
SQ Sequence 1255 AA;
Query Match 98.8%; Score 6720; DB 22; Length 1255;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1242; Conservative 2; Mismatches 11; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLRLPASPETHLDMLRLHYQGCQVVGNNL 60
Db 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLRLPASPETHLDMLRLHYQGCQVVGNNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQGVYLIHQVQVPLQRLIRVRGTQLPEDNALVALVDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQGVYLIHQVQVPLQRLIRVRGTQLPEDNALVALVDNG 120
QY 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGVLIQRPOLQVQDTILWKDIFHKNNOLA 180
Db 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGVLIQRPOLQVQDTILWKDIFHKNNOLA 180
QY 181 LTLIDNTRSFACHPCSPCKSGKSGESSEDQCSLTRVTCAGGCARCKGPLETDCCHQC 240
Db 181 LTLIDNTRSFACHPCSPCKSGKSGESSEDQCSLTRVTCAGGCARCKGPLETDCCHQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVQYIKANSKFGITELRITFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVQYIKANSKFGITELRITFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHMQEVTASDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHMQEVTASDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
QY 361 IOEFAGCKKIFGSLAPLPESEFDGDPASNTAPLQBPQLQVFETLEETGYLYISAWPDSLP 420
Db 361 IOEFAGCKKIFGSLAPLPESEFDGDPASNTAPLQBPQLQVFETLEETGYLYISAWPDSLP 420
QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLRSLRELGSGLALIHNNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLRSLRELGSGLALIHNNTHLCFVHTV 480
QY 481 PWDQLFRPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPGPTQVCNCSQFLRGQSC 540
Db 481 PWDQLFRPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPGPTQVCNCSQFLRGQSC 540
QY 541 VEECRVLOGLPREYVYVNAHCLPCHPECOFQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
Db 541 VEECRVLOGLPREYVYVNAHCLPCHPECOFQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
QY 601 PSGVKPDLISYMPDWKPPDEEGACQPCPNCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660
Db 601 PSGVKPDLISYMPDWKPPDEEGACQPCPNCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660
QY 661 ILLVVVLGVVFGILIKRROOKIRKYMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRROOKIRKYMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720

721 RKVKLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTSPKANKELIDYAVMAGVSP 780
 721 RKVKLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTSPKANKELIDYAVMAGVSP 780
 781 YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRRGLSGDQLLNCWQIAKMSYLEDVR 840
 781 YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRRGLSGDQLLNCWQIAKMSYLEDVR 840
 841 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900
 841 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900
 901 HQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYIMVKWM 960
 901 HQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYIMVKWM 960
 961 IDSECRPRRELVSFSESRMARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDMDGLVDA 1020
 961 IDSECRPRRELVSFSESRMARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDMDGLVDA 1020
 1021 EYLVPOQGFPCPDPAAGAGWVHRRSSSTRSGGDLTLGLEPSEERAPRPLAPSEG 1080
 1021 EYLVPOQGFPCPDPAAGAGWVHRRSSSTRSGGDLTLGLEPSEERAPRPLAPSEG 1080
 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAPGGAVENPEYLTQ 1200
 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAPGGAVENPEYLTQ 1200
 1201 GGAPOHPHPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
 1201 GGAPOHPHPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 3

AAB60167
 ID AAB60167 standard; Protein; 1255 AA.

XX AC AAB60167;

XX 03-APR-2001 (first entry)

XX HRP2 transgene plasmid construct encoded protein.

XX Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
 antibody.

XX Homo sapiens.

XX Synthetic.

XX WO200100244-A2

XX 04-JAN-2001

XX 23-JUN-2000; 2000WO-US17229.

XX 25-JUN-1999; 99US-0141316.

XX 16-MAR-2000; 2000US-0189844.

XX (GETH) GENENTECH INC.

XX Erickson S, Schwall R;

XX WPI; 2001-061962/07.

XX N-PSDB; AAF24297.

XX Treating tumors, particularly breast cancers, which overexpress an ErbB
 PT receptor and does not respond to an anti-ErbB antibody, comprises
 PT conjugating the antibody to a maytansinoid -

XX

PS Example 3; Fig 4; 92pp; English.

XX The present invention provides a method of treating cancer by
 CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
 CC particular, the antibody is directed against ErbB2 (also known as HER2
 CC and p185neu). The method is particularly useful in the treatment of
 CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
 CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.

XX Sequence 1255 AA;

Query Match 98.8%; Score 6720; DB 22; Length 1255;

Best Local Similarity 99.0%; Pred. No. 0;

Matches 1242; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASSTVCTGTDMLKRLPASPETHLDMLRHLYQGCVVQGNL 60
 DB 1 MELAALCRWGLLLALLPPGAASSTVCTGTDMLKRLPASPETHLDMLRHLYQGCVVQGNL 60
 QY 61 ELYLPTNASLSFLQDIQEVQGVLIHAHQVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
 DB 61 ELYLPTNASLSFLQDIQEVQGVLIHAHQVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
 QY 121 DPLNNTTPTVTGASPGGLRELQRLSLEILKGGVLIQORNQPCYQDTILWKDIFHKNNQLA 180
 DB 121 DPLNNTTPTVTGASPGGLRELQRLSLEILKGGVLIQORNQPCYQDTILWKDIFHKNNQLA 180
 QY 181 LTLIDNRSRACHPCSPMKCSRCWGESSEDCQSLTRTVCAAGCARCKGPLPDDCCEQC 240
 DB 181 LTLIDNRSRACHPCSPMKCSRCWGESSEDCQSLTRTVCAAGCARCKGPLPDDCCEQC 240
 QY 241 AAGCTGPKHSDCLACLFHNSHGICELHCPALVQYIKANSKFIGITELRVTFFGASCVTACP 300
 DB 241 AAGCTGPKHSDCLACLFHNSHGICELHCPALVQYIKANSKFIGITELRVTFFGASCVTACP 300
 QY 301 YNYLSTDVGSCTLVCPHLNHQEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
 DB 301 YNYLSTDVGSCTLVCPHLNHQEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
 QY 361 IQEFAGCKKI FGSFLAPLSPESFGDPSASNTAPLOPEQLQVFETLEEITGVLYISAWPDSL 420
 DB 361 IQEFAGCKKI FGSFLAPLSPESFGDPSASNTAPLOPEQLQVFETLEEITGVLYISAWPDSL 420
 QY 421 DLSVFONLQVIRGRIILHNGAYSITLQGLIGISWGLRSRLRELGLSLALIHNNHLCFVHTV 480
 DB 421 DLSVFONLQVIRGRIILHNGAYSITLQGLIGISWGLRSRLRELGLSLALIHNNHLCFVHTV 480
 QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGTQCVCNCSQFLRGQEC 540
 DB 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGTQCVCNCSQFLRGQEC 540
 QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVCANHYKDPFPCVARC 600
 DB 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVCANHYKDPFPCVARC 600
 QY 601 PSGVKPDLSPYMPIWKPDEEGACOPCPINCTHSCVDLDDKGCPEARSPALTSIVSAVVG 660
 DB 601 PSGVKPDLSPYMPIWKPDEEGACOPCPINCTHSCVDLDDKGCPEARSPALTSIVSAVVG 660
 QY 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
 DB 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
 QY 721 RKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTSPKANKELIDYAVMAGVSP 780
 DB 721 RKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTSPKANKELIDYAVMAGVSP 780
 QY 781 YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRRGLSGDQLLNCWQIAKMSYLEDVR 840
 DB 781 YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRRGLSGDQLLNCWQIAKMSYLEDVR 840
 QY 841 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900

Db 841 LVHRDLARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKMKMALESILRRFT 900
Qy 901 HQSDVWSYGVTVWELMTFGAKPDGIPAREIPDLLEKGERLPQPPICTIDVYIMVKWM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPDGIPAREIPDLLEKGERLPQPPICTIDVYIMVKWM 960
Qy 961 IDECRPRFRELVSFMRDQRFVVIQNEDLGPASPLDSTFYRSLLEDMDGLVDA 1020
Db 961 IDECRPRFRELVSFMRDQRFVVIQNEDLGPASPLDSTFYRSLLEDMDGLVDA 1020
Qy 1021 EYLVPQGFCCPDPAAGWVHRRHSSTSGGGDLTLGLEPSEERAPPLAPSEG 1080
Db 1021 EYLVPQGFCCPDPAAGWVHRRHSSTSGGGDLTLGLEPSEERAPPLAPSEG 1080
Qy 1081 AGSDVDFDGLGMAAKQLSLPHTDPSPLQRYSEDPTVPLPSETDGYVAFELTCSPOPEYV 1140
Db 1081 AGSDVDFDGLGMAAKQLSLPHTDPSPLQRYSEDPTVPLPSETDGYVAFELTCSPOPEYV 1140
Qy 1141 NOPDVRPQSPREGPLPAARPAATLAKTLSPGKNGVVKDVFAPGGAVENPEYLTPO 1200
Db 1141 NOPDVRPQSPREGPLPAARPAATLAKTLSPGKNGVVKDVFAPGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGPTAENPEYLGIDVPV 1255
Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGPTAENPEYLGIDVPV 1255

RESULT 4

AAU74545
ID AAU74545 standard; Protein; 1255 AA.

AC AAU74545;

XX 23-APR-2002 (first entry)

XX Human HER2/ (ErbB2) polypeptide.

XX Human; HER2; ErbB; epidermal growth factor receptor; receptor;
KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
KW glial disorder; astrocytal disorder; hypothalamic disorder;
KW glandular disorder; macrophagal disorder; epithelial disorder;
KW stromal disorder; blastocoealic disorder; inflammatory disorder;
KW angiogenic disorder; immunological disorder.

XX Homo sapiens.

XX US2002001587-A1.

XX 03-JAN-2002.

XX 16-MAR-2001; 2001US-0811123.

XX 16-MAR-2000; 2000US-189844P.

PR 05-OCT-2000; 2000US-238327P.

XX (ERIC/) ERICKSON S.

PA (SCHW/) SCHWALL R.

PA (SLIW/) SLIWOWSKI M.

XX Erickson S, Schwall R, Sliwowski M;

XX WPI; 2002-163686/21.

DR N-PSDB; ABK14058.

XX Treating tumour characterised by overexpression of epidermal growth
PT factor receptor, ErbB or cancer in mammal, comprises administering
PT anti-ErbB antibody-maytansinoid conjugate to the mammal

XX Example 3; Fig 7; 93pp; English.

XX The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor
CC receptor (ErbB) and does not respond or responds poorly, to treatment
CC with an anti-ErbB antibody, comprising administering to the mammal an
CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
CC prostate and bladder, preferably breast cancer. The breast cancer is a
CC metastatic breast cancer or an aggressive form of metastatic breast
CC cancer which overexpresses ErbB2. The method is also useful for treating
CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
CC epithelial, stromal, blastocoealic, inflammatory, angiogenic and
CC immunological disorders. This sequence represents the human HER2 (ErbB2)
CC polypeptide of the invention.

XX Sequence 1255 AA;

Query Match 98.8%; Score 6720; DB 23; Length 1255;

Best Local Similarity 99.0%; Pred. No. 0;

Matches 1242; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLPPGAASDVCTGTDMLRLPASPEHDLMLRLHYQGCVVQGNL 60

Db 1 MELAALCRWGLLLALLPPGAASDVCTGTDMLRLPASPEHDLMLRLHYQGCVVQGNL 60

Qy 61 ELYLPTNASLSFLQDIQEVQGVYLIHQNVRQVPLQRLIRVGTOLFDENVALAVLDNG 120

Db 61 ELYLPTNASLSFLQDIQEVQGVYLIHQNVRQVPLQRLIRVGTOLFDENVALAVLDNG 120

Qy 121 DPLNNTPTVTGASPGGLRELQSLTEILKGGVLTORNPOLCYODTILKNDIFHKNNOLA 180

Db 121 DPLNNTPTVTGASPGGLRELQSLTEILKGGVLTORNPOLCYODTILKNDIFHKNNOLA 180

Qy 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVTCAGGCARCKGPLTDCCHCEQC 240

Db 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVTCAGGCARCKGPLTDCCHCEQC 240

Qy 241 AAGCTGPKHSIDCLAHFNHSGICELHCPALVYIKANSKFITELRTVFGASCVTACP 300

Db 241 AAGCTGPKHSIDCLAHFNHSGICELHCPALVYIKANSKFITELRTVFGASCVTACP 300

Qy 301 YNVLSTDVGSCTLVCPHNOEVTABDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360

Db 301 YNVLSTDVGSCTLVCPHNOEVTABDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360

Qy 361 IQEFAGCKKIFGSLAPLPESFGDPPASNTAPLQPEQLQVFETLEETIGLYISAWPDSLP 420

Db 361 IQEFAGCKKIFGSLAPLPESFGDPPASNTAPLQPEQLQVFETLEETIGLYISAWPDSLP 420

Qy 421 DLSVFONLQVIRGRILHNGAYSILTQGLGISWLGRLSRLRELGLALIHNNTHLCFVHTV 480

Db 421 DLSVFONLQVIRGRILHNGAYSILTQGLGISWLGRLSRLRELGLALIHNNTHLCFVHTV 480

Qy 481 PWDQLFRNPHOALLHTANRPEDECYEGELACHQLCARGHCWGPPTQCNCQSLRGQEC 540

Db 481 PWDQLFRNPHOALLHTANRPEDECYEGELACHQLCARGHCWGPPTQCNCQSLRGQEC 540

Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600

Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600

Qy 601 PSGVKPDLISYMPIWKFPDDEGACQPCPNCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660

Db 601 PSGVKPDLISYMPIWKFPDDEGACQPCPNCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660

Qy 661 ILLVVLGVVFGILIKRROOKIRKVTMRLLQETELVEPLTPSGAMPNOAQRILKETEL 720

Db 661 ILLVVLGVVFGILIKRROOKIRKVTMRLLQETELVEPLTPSGAMPNOAQRILKETEL 720

Qy 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVALKURENTSPKANKEIILDEAYVMAGVGP 780

Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVALKURENTSPKANKEIILDEAYVMAGVGP 780

Qy 781 YVSRLLIGICLTSTVOLVTLMPYGCILLDHVRENRRGLSGQDLLNMCQIAGKMSYLEVDVR 840

Db 781 YVSRLLGICLTSTVQLVTQMPYGLDHRVNRGRGLSGDQLLNWCQIAKMSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLDIDETEHADGGKVPKIMMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLDIDETEHADGGKVPKIMMALESILRRFT 900
Qy 901 HQSDVNSYGVTVWELMTFGAKPDGIPAREIPDLLEKGERLPQPPICITIDVYIMVWKWM 960
Db 901 HQSDVNSYGVTVWELMTFGAKPDGIPAREIPDLLEKGERLPQPPICITIDVYIMVWKWM 960
Qy 961 IDSECRPRRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDEDDWGLDVA 1020
Db 961 IDSECRPRRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDEDDWGLDVA 1020
Qy 1021 EYVLVPOQGFCDPAPGAGVHHRSSSTRSGGDLTLGLEPSEERAPRPLAPSEG 1080
Db 1021 EYVLVPOQGFCDPAPGAGVHHRSSSTRSGGDLTLGLEPSEERAPRPLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAGLQSLPTHDPSLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGLQSLPTHDPSLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Qy 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTQ 1200
Db 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTQ 1200
Qy 1201 GGAAPQHPHPPAPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255
Db 1201 GGAAPQHPHPPAPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255

RESULT 5

AAW01111 standard; Protein; 1255 AA.

AC AAW01111;
01-JAN-1997 (first entry)
DE HER-2/neu protein.

XX HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW breast cancer; ovary cancer; colon cancer; lung cancer;
KW prostate cancer; immunisation; tumour; vaccine; vector.
OS Homo sapiens.

XX Key Location/Qualifiers
FH Domain 676..1255
FT /label= Intracellular domain
FT /note= "claimed domain, useful for immunisation"

XX 28-MAR-1996; 96WO-US01689.
XX 31-MAR-1995; 95US-0414417.
XX (UNIW) UNIV WASHINGTON.

XX Cheever MA, Disis ML;
XX WPI; 1996-455361/45.
XX N-PSDB; AAT40739.

XX DNA encoding HER-2-neu poly-peptide(s) - used for prevention or
XX treatment of malignancies with which the HER-2/neu oncogene is
XX associated
XX Claim 2; Page 56-61; 71pp; English.

XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transfected host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.

XX SQ Sequence 1255 AA;

Query Match 98.7%; Score 6714; DB 17; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLALLPAGAASTQVCTGTDMLRLPASPTHLDMLRHLYQGCVVQGNL 60
Db 1 MELAALCRWGLLALLPAGAASTQVCTGTDMLRLPASPTHLDMLRHLYQGCVVQGNL 60
Qy 61 ELTYLPTNASLSFLODIQEVQVYLIHNVQVPLQRLRIVRGTLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLODIQEVQVYLIHNVQVPLQRLRIVRGTLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNOLA 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNOLA 180
Qy 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVTCAGGCARCKGPLPTDCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVTCAGGCARCKGPLPTDCHEQC 240
Qy 241 AAGCTGPKSDCLACLFHNSGICELHCPALVQYIKANSKFICITEIRTFGASCVTACP 300
Db 241 AAGCTGPKSDCLACLFHNSGICELHCPALVQYIKANSKFICITEIRTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPHMQEVTAEQDTCRCKSPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHMQEVTAEQDTCRCKSPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLVFTELEEITGYLISAMPDLSL 420
Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLVFTELEEITGYLISAMPDLSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLISWLGSLRELGSGLALIIHHNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLISWLGSLRELGSGLALIIHHNTHLCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQCVCNCSQFLRGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQCVCNCSQFLRGQEC 540
Qy 541 VEECRVLOGLPREYVNAHCLPCHPECQPNQSVTCFGEADQCVACAHYKDPFPCVARC 600
Db 541 VEECRVLOGLPREYVNAHCLPCHPECQPNQSVTCFGEADQCVACAHYKDPFPCVARC 600
Qy 601 PSGVKPDLSPYMIWKFPPDEGACQPCPINCTHSCVDLDDKGCPEARASPLTSIVSAVWG 660
Db 601 PSGVKPDLSPYMIWKFPPDEGACQPCPINCTHSCVDLDDKGCPEARASPLTSIVSAVWG 660
Qy 661 ILLVVVLGVVFGILIKRROOKIRKYIMRRLLOETELVEPLTPSGAMPNOAQMILKETEL 720
Db 661 ILLVVVLGVVFGILIKRROOKIRKYIMRRLLOETELVEPLTPSGAMPNOAQMILKETEL 720
Qy 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVSP 780
Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVSP 780
Qy 781 YVSRLLGICLTSTVQLVTQMPYGLDHRVNRGRGLSGDQLLNWCQIAKMSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTQMPYGLDHRVNRGRGLSGDQLLNWCQIAKMSYLEDVR 840

QY 841 LVHRDLAARNVLKSPNHHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRFT 900
 DB 841 LVHRDLAARNVLKSPNHHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRFT 900
 QY 901 HQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYMWKWM 960
 DB 901 HQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYMWKWM 960
 QY 961 IDSECRPRFRELVSERMRARDQRFVVIQNEDLGPASPLDSTFYRSLLDDDDMDGLVDA 1020
 DB 961 IDSECRPRFRELVSERMRARDQRFVVIQNEDLGPASPLDSTFYRSLLDDDDMDGLVDA 1020
 QY 1021 EYLVVCOQGFCDPAPAGAGMVHRRSSSTRSGGDLTLGLEPSEERAPRLAPSEG 1080
 DB 1021 EYLVVCOQGFCDPAPAGAGMVHRRSSSTRSGGDLTLGLEPSEERAPRLAPSEG 1080
 QY 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLORYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
 DB 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLORYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
 QY 1141 NQDVRPQPSPREGPLPAARPAAGATILERAKTILSPGKNGVVKDVFAFGGAVENPEYLTPO 1200
 DB 1141 NQDVRPQPSPREGPLPAARPAAGATILERAKTILSPGKNGVVKDVFAFGGAVENPEYLTPO 1200
 QY 1201 GGAAPQHPHPPAPSPAFDNLVYWDQPPPERGAPPSTFKGTPPTAENPEYLGLDVVPV 1255
 DB 1201 GGAAPQHPHPPAPSPAFDNLVYWDQPPPERGAPPSTFKGTPPTAENPEYLGLDVVPV 1255

RESULT 6

AAW92406
 ID AAW92406 standard; Protein; 1255 AA.

XX AAW92406;
 AC
 XX
 DT 21-APR-1999 (first entry)
 XX
 DE Human HER-2/neu oncogene protein.
 XX
 KW HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
 KW malignancy; treatment; tumour.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Region 676..1255
 FT /note= "region which elicits immune response"

XX US5869445-A.

XX 09-FEB-1999.

XX 01-APR-1996; 96US-0625101.

XX 01-APR-1996; 96US-0625101.

PR 17-MAR-1993; 93US-0033644.

PR 12-AUG-1993; 93US-0106112.

PR 31-MAR-1995; 95US-0414417.

XX (UNIW) UNIV WASHINGTON.

XX Cheever MA, Disis ML;

XX WPI; 1999-152835/13.

DR N-PSDB; AAX01912.

XX Use of HER-2/neu polypeptides - for eliciting an immune response to

PT an HER-2/neu associated malignancy, particularly for treating or

PT preventing tumours

XX Claim 3; Column 31-38; 26pp; English.

PS This sequence represents the human HER-2/neu oncogene protein. A fragment

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC of this protein is used in a method for eliciting or enhancing an immune
 CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
 CC B cells to produce an immune response to the HER-2/neu protein. The
 CC method can be used for immunisation against a malignancy in which the
 CC HER-2/neu oncogene is associated and in the treatment of an existing
 CC tumour, or to prevent tumour occurrence or reoccurrence.

XX Sequence 1255 AA;

Query Match 98.7%; Score 6714; DB 20; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;

Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLRPASPTHLDMLRLHYQGCVVQGNL 60

DB 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLRPASPTHLDMLRLHYQGCVVQGNL 60

QY 61 ELYLPTNASLSFLQDIQEVQGVLIHQNVRQVPLQRLIRVGTOLFDENYALAVLDNG 120

DB 61 ELYLPTNASLSFLQDIQEVQGVLIHQNVRQVPLQRLIRVGTOLFDENYALAVLDNG 120

QY 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKKNOLA 180

DB 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKKNOLA 180

QY 181 LTLIDTNRSFACHPCSPCKSGRCWGESSEDCQSLTRITVCAGGCARCKGPLPTDCCHQC 240

DB 181 LTLIDTNRSFACHPCSPCKSGRCWGESSEDCQSLTRITVCAGGCARCKGPLPTDCCHQC 240

QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVQYIKANSKFTIGITELRTVFGASCVTACP 300

DB 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVQYIKANSKFTIGITELRTVFGASCVTACP 300

QY 301 YNYLSTDVGSCTLVCPHMQEVTAEQGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360

DB 301 YNYLSTDVGSCTLVCPHMQEVTAEQGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360

QY 361 IOEFAGCKKIFGSLAFPLPESFDGDPASNTAPLQPEQLQVFETLEETIGLYISAWPDSLP 420

DB 361 IOEFAGCKKIFGSLAFPLPESFDGDPASNTAPLQPEQLQVFETLEETIGLYISAWPDSLP 420

QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLRSLRELGLSLALIHNNHLCFVHTV 480

DB 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLRSLRELGLSLALIHNNHLCFVHTV 480

QY 481 PNDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSQFLRGQEC 540

DB 481 PNDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSQFLRGQEC 540

QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFFCVARC 600

DB 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFFCVARC 600

QY 601 PSGVKPDLISYMPIWKFPPDEEGACQPCPNCTHSCVDLDDKCPAERASPLTSISAVVG 660

DB 601 PSGVKPDLISYMPIWKFPPDEEGACQPCPNCTHSCVDLDDKCPAERASPLTSISAVVG 660

QY 661 ILLVVLGVVFGILIKRROOKIRKVTMRLLQETELVEPLTPSGAMPNQAORILKETEL 720

DB 661 ILLVVLGVVFGILIKRROOKIRKVTMRLLQETELVEPLTPSGAMPNQAORILKETEL 720

QY 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVRENTSPKANKEILDEAYMAGVGP 780

DB 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVRENTSPKANKEILDEAYMAGVGP 780

QY 781 YVSRLLGICLTSTVOLVTQMPYGCCLLDHVRNRLGSLQDQLLNMCQIAKMSYLEDVYR 840

DB 781 YVSRLLGICLTSTVOLVTQMPYGCCLLDHVRNRLGSLQDQLLNMCQIAKMSYLEDVYR 840

QY 841 LVHRDLAARNVLKSPNHHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRFT 900

DB 841 LVHRDLAARNVLKSPNHHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRFT 900

QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKWM 960
 DB |||||
 QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKWM 960
 DB |||||
 QY 961 IDSECRPRFRELVSERFMRARDPQRFVVIQNEDLGPASPLDSTFYRSLLDEDDMDGLVDA 1020
 DB |||||
 QY 961 IDSECRPRFRELVSERFMRARDPQRFVVIQNEDLGPASPLDSTFYRSLLDEDDMDGLVDA 1020
 DB |||||
 QY 1021 EYLVVQOQFFCDPAPGAGCMVHRRSSSTRSGGDLTLGLEPSEERAPRSLAPSEG 1080
 DB |||||
 QY 1021 EYLVVQOQFFCDPAPGAGCMVHRRSSSTRSGGDLTLGLEPSEERAPRSLAPSEG 1080
 DB |||||
 QY 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSTGTVAPLTCSPQPEYV 1140
 DB |||||
 QY 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSTGTVAPLTCSPQPEYV 1140
 DB |||||
 QY 1141 NQDVRPQPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVFAFGGAVENPEYLTPO 1200
 DB |||||
 QY 1141 NQDVRPQPSPREGPLPAARPAAGATLERPKTILSPGKNGVYKDVFAFGGAVENPEYLTPO 1200
 DB |||||
 QY 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERCAPPSTFKGTPTAENPEYLGLDVVPV 1255
 DB |||||
 QY 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERCAPPSTFKGTPTAENPEYLGLDVVPV 1255
 DB |||||

RESULT 7

AAB21198 standard; protein; 1255 AA.

QY AAB21198;

QY 12-JAN-2001 (first entry)

QY Human HER-2/neu protein.

QY Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;

QY breast cancer; prostate cancer; ovarian cancer; lung cancer;

QY colon cancer.

QY Homo sapiens.

QY WO200044899-A1.

QY 03-AUG-2000.

QY 28-JAN-2000; 2000WO-US02164.

QY 29-JAN-1999; 99US-0117976.

QY (CORI-) CORIXA CORP.

QY (SMIK) SMITHKLINE BEECHAM.

QY Cheever MA, Gheysen D;

QY WPI; 2000-505976/45.

QY N-PSDB; AAA89736.

QY HER-2/neu extracellular domain/phosphorylation domain fusion proteins

QY useful for vaccinating against breast, ovarian, colon, lung and

QY prostate cancers -

QY Claim 52; Fig 7; 128pp; English.

QY The present sequence is the human HER-2/neu protein. It is a member of

QY the tyrosine kinase family of receptor-like glycoproteins and shows

QY homology to the epidermal growth factor receptor (EGFR). It probably

QY plays a part in cell growth and/or differentiation. The HER-2/neu

QY gene is an oncogene. An HER-2/neu fusion protein comprising a

QY HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation

QY domain may be used to treat or prevent cancer by eliciting or

QY enhancing an immune response to the HER-2/neu protein. It may be used

QY to treat malignancies such as breast, ovarian, colon, lung and

QY prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.
 XX
 SQ Sequence 1255 AA;
 Query Match 98.7%; Score 6714; DB 21; Length 1255;
 Best Local Similarity 98.8%; Pred. No. 0;
 Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;
 QY 1 MELAALCRWGLLALLPPGAASVCTGTDMKRLRPASPTHLDMLRHLYQCCVVOGNL 60
 DB |||||
 QY 1 MELAALCRWGLLALLPPGAASVCTGTGTDMKRLRPASPTHLDMLRHLYQCCVVOGNL 60
 DB |||||
 QY 61 ELTYLPTNASLSFLQDIQEVQGVLIHMQVPLQRLRIVRGTQLFEDNTALAVLDNG 120
 DB |||||
 QY 61 ELTYLPTNASLSFLQDIQEVQGVLIHMQVPLQRLRIVRGTQLFEDNTALAVLDNG 120
 DB |||||
 QY 121 DPLNNTTPTVGASPGGLRELRLSLEILKGGVLIQORNPOLCYQDTILWKDIFHKKNOLA 180
 DB |||||
 QY 121 DPLNNTTPTVGASPGGLRELRLSLEILKGGVLIQORNPOLCYQDTILWKDIFHKKNOLA 180
 DB |||||
 QY 181 LTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSLTRTYCAGGCARCKGPLPTDCCHEQC 240
 DB |||||
 QY 181 LTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSLTRTYCAGGCARCKGPLPTDCCHEQC 240
 DB |||||
 QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVQYIKANSKFITGITELRITFGASCVTACP 300
 DB |||||
 QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVQYIKANSKFITGITELRITFGASCVTACP 300
 DB |||||
 QY 301 YNYLSTDVSGCTLVCPHLNHOEVTAEQTCRCKSKPCARVCYGLGMEHLREVRVTSAN 360
 DB |||||
 QY 301 YNYLSTDVSGCTLVCPHLNHOEVTAEQTCRCKSKPCARVCYGLGMEHLREVRVTSAN 360
 DB |||||
 QY 361 IOEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPEQLQVFETLEETITGLYISAWPDSLP 420
 DB |||||
 QY 361 IOEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPEQLQVFETLEETITGLYISAWPDSLP 420
 DB |||||
 QY 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGSIWGLSLRELGLSLALIHNNTHLCFVHTV 480
 DB |||||
 QY 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGSIWGLSLRELGLSLALIHNNTHLCFVHTV 480
 DB |||||
 QY 481 PWDQLFRNPQALLTANRPEDECVGEGLACHQLCARGHCWGPGTQCVNCSOFLRGQEC 540
 DB |||||
 QY 481 PWDQLFRNPQALLTANRPEDECVGEGLACHQLCARGHCWGPGTQCVNCSOFLRGQEC 540
 DB |||||
 QY 541 VEECRVLQGLPREYNNARHCLPCHPECQONGSVTCFGEADQCVACAHYKOPPPFCVARC 600
 DB |||||
 QY 541 VEECRVLQGLPREYNNARHCLPCHPECQONGSVTCFGEADQCVACAHYKOPPPFCVARC 600
 DB |||||
 QY 601 PSGVKPDLISYMPIWKPFPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660
 DB |||||
 QY 601 PSGVKPDLISYMPIWKPFPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660
 DB |||||
 QY 661 ILLVVVLGVVFGILIKRROOKIRKYTMRELLQETELVEPLTPSGAMPNOAQMRILKETEL 720
 DB |||||
 QY 661 ILLVVVLGVVFGILIKRROOKIRKYTMRELLQETELVEPLTPSGAMPNOAQMRILKETEL 720
 DB |||||
 QY 721 RKVKVLGSAFGTVYKGIWIPGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGP 780
 DB |||||
 QY 721 RKVKVLGSAFGTVYKGIWIPGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGP 780
 DB |||||
 QY 781 YVSRLLGICLTSTVOLVTLQMPYGLLDHVRNRRGLSGQDLNLCWCMQIAGMSYLEDVR 840
 DB |||||
 QY 781 YVSRLLGICLTSTVOLVTLQMPYGLLDHVRNRRGLSGQDLNLCWCMQIAGMSYLEDVR 840
 DB |||||
 QY 841 LVHRDLAARNVLKSPNNHVKITDFGLARLLDIDETEHADGGKVPKKNWALSILRRRFT 900
 DB |||||
 QY 841 LVHRDLAARNVLKSPNNHVKITDFGLARLLDIDETEHADGGKVPKKNWALSILRRRFT 900
 DB |||||
 QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKWM 960
 DB |||||
 QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKWM 960
 DB |||||
 QY 961 IDSECRPRFRELVSERFMRARDPQRFVVIQNEDLGPASPLDSTFYRSLLDEDDMDGLVDA 1020
 DB |||||

Db 961 IDSECRPRFRELVSERFARMARDQRFVVIQNEIDLGASPASLDSTFYRSLLEDMDGLVDA 1020
QY 1021 EYLVPQOQFFCPDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Db 1021 EYLVPQOQFFCPDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
QY 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEYV 1140
QY 1141 NOPDVQPQPSREGPLPAARPAAGATLERAKTILSPGKNGVVVDVFAFGGAVENPEYLTPO 1200
Db 1141 NOPDVQPQPSREGPLPAARPAAGATLERAKTILSPGKNGVVVDVFAFGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGPTTAENPEYLGLDVVP 1255
Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGPTTAENPEYLGLDVVP 1255

RESULT 8

AAy84780
ID AAY84780 standard; Protein; 1255 AA.
XX
AC AAY84780;
XX
DT 08-AUG-2000 (first entry)
XX
DE Amino acid sequence of the SPLICE erbB-2 receptor protein.
XX
KW SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;
KW tumor cell proliferation; tissue degeneration; arthropathy;
KW bone resorption; inflammatory disease; degenerative disorder;
KW wound healing.
XX
OS Homo sapiens.
XX
FN WO200020579-A1.
XX
PD 13-APR-2000.
XX
PF 01-OCT-1999; 99WO-CA00912.
XX
PR 02-OCT-1998; 98US-0165192.
XX
PA (UYMC-) UNIV MCMASTER.
XX
PI Muller WJ, Siegel PM;
XX
DR WPI; 2000-303768/26.
XX
DR N-PSDB; AAA14812.

XX
PT Nucleic acid encoding an erbB 2 receptor protein designated SPLICE
PT erbB-2, inhibitors of the protein are useful for treatment of cancer -
PS
PS Claim 3; Fig 2; 60pp; English.

XX
XX The present sequence represents a SPLICE erbB-2 receptor protein. The
CC protein has an in-frame deletion of 16 amino acids, 2 of which are
CC conserved cysteine residues, compared to the unspliced protein. The
CC erbB-2 polynucleotide is used to construct probes for detecting
CC disorders of cell transformation such as cancer. Antibodies to the
CC protein may be used to detect SPLICE erbB-2 in a sample. Agents
CC (e.g. antisense oligonucleotides) which inhibit the expression of
CC SPLICE erbB-2 are useful for reducing tumor cell proliferation and
CC treating cancer. Substances which stimulate SPLICE erbB-2 are useful
CC for treating conditions involving damaged cells including conditions
CC in which degeneration of tissue occurs, such as arthropathy, bone
CC resorption, inflammatory diseases, degenerative disorders of the
CC central nervous system and wound healing.

XX Sequence 1255 AA;

Query Match 98.7%; Score 6714; DB 21; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;
QY 1 MELAALCRWELLALLPPCAASTOCTGDMKLRIPASPTHLDMLRLHYQGCQVQGNL 60
Db 1 MELAALCRWELLALLPPCAASTOCTGDMKLRIPASPTHLDMLRLHYQGCQVQGNL 60
QY 61 ELTYLPTNASLSFLQDIEVOGVYLIHNRQVPLQRLRIVRGTOLFFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIEVOGVYLIHNRQVPLQRLRIVRGTOLFFEDNYALAVLDNG 120
QY 121 DPLNNTPTVTGASPGGLRELQRLSTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNOLA 180
Db 121 DPLNNTPTVTGASPGGLRELQRLSTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNOLA 180
QY 181 LTLIDNTRSRACHPCSPCKGRCWGSESDCQSLTRTVACGCAKCKGKPLPTDCCHQEC 240
Db 181 LTLIDNTRSRACHPCSPCKGRCWGSESDCQSLTRTVACGCAKCKGKPLPTDCCHQEC 240
QY 241 AAGCTGPKHSDCLACILHFNHSGICELHCPALVQYIKANSKFICITELRYTTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACILHFNHSGICELHCPALVYNTDTFESMPNPSGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHNRQVTAEDGTORCEKCKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHNRQVTAEDGTORCEKCKPCARVCYGLGMEHLREVRVTSAN 360
QY 361 IOEFGAGCKKIFGSLAPLPESFDGDPASNTAPLOPOLQVTFLEETIGVLYISAWPDSLP 420
Db 361 IOEFGAGCKKIFGSLAPLPESFDGDPASNTAPLOPOLQVTFLEETIGVLYISAWPDSLP 420
QY 421 DLSVFONLQVIRGRILHNGAYSILTLQGLIGISWGLSRLSRELGSGLALIHNNHLCFVHTV 480
Db 421 DLSVFONLQVIRGRILHNGAYSILTLQGLIGISWGLSRLSRELGSGLALIHNNHLCFVHTV 480
QY 481 PWDQLFRNPHQALLTANRPEDECVGEGIAHOLCARGHCWGPGPTQCVNCSOFLRGQEC 540
Db 481 PWDQLFRNPHQALLTANRPEDECVGEGIAHOLCARGHCWGPGPTQCVNCSOFLRGQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPECOQNGSVTCFGEADQCVACAHYKOPPCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECOQNGSVTCFGEADQCVACAHYKOPPCVARC 600
QY 601 PSGVKPDLSYMPIWKFPPDEGACQPCINCTHSCVDLDDKGCAPAEORASPLTSIVSAVVG 660
Db 601 PSGVKPDLSYMPIWKFPPDEGACQPCINCTHSCVDLDDKGCAPAEORASPLTSIVSAVVG 660
QY 661 ILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNOAQMRLKETEL 720
Db 661 ILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNOAQMRLKETEL 720
QY 721 RKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYMAGVGP 780
Db 721 RKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYMAGVGP 780
QY 781 YVSRLLGICLTSTVOLVTOLMPYGLLDHVRNRRGLSGDQLLNCWQIAKMSYLEDVYR 840
Db 781 YVSRLLGICLTSTVOLVTOLMPYGLLDHVRNRRGLSGDQLLNCWQIAKMSYLEDVYR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRRFT 900
QY 901 HQSDVMSYGVTVWELMTFGAKYDGPAREIPDLLEKGERLPOPPICITIDVTMVKWM 960
Db 901 HQSDVMSYGVTVWELMTFGAKYDGPAREIPDLLEKGERLPOPPICITIDVTMVKWM 960
QY 961 IDSECRPRFRELVSERFARMARDQRFVVIQNEIDLGASPASLDSTFYRSLLEDMDGLVDA 1020
Db 961 IDSECRPRFRELVSERFARMARDQRFVVIQNEIDLGASPASLDSTFYRSLLEDMDGLVDA 1020
QY 1021 EYLVPQOQFFCPDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080

Db 1021 EEVLVPOQGFCDPAPAGAGWVHRRSSSTRSGGDLTLGLEPSEERAPSPAPSEG 1080
QY 1081 AGSDVFDGDLGMAKGLQSLPTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAKGLQSLPTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
QY 1141 NQPDVRQPSPREGPLPAARPAATLERAKTILSPGKGVVVDVAFGGAVENPEYLTPO 1200
Db 1141 NQPDVRQPSPREGPLPAARPAATLERAKTILSPGKGVVVDVAFGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPPPAFSPADNLYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
Db 1201 GGAAPQHPPPAFSPADNLYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 9
AAB85458
ID AAB85458 standard; Protein; 1255 AA.
XX AAB85458;
XX AC AAB85458;
XX DE 25-SEP-2001 (first entry)
XX Human HER-2/neu protein.
XX Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
XX oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX Homo sapiens.
XX WO200153463-A2.
XX 26-JUL-2001.
XX 19-JAN-2001; 2001WO-US01850.
XX 21-JAN-2000; 2000US-0177545.
XX (CORI-) CORIXA CORP.
XX Cheever MA, Hand-Zimmermann S;
XX WPI; 2001-476112/51.
XX N-PSDB; AAH23392.
XX New antigen-presenting cells, useful as vaccines for eliciting or
XX enhancing an immune response to HER-2/neu protein, particularly useful
XX for treating or preventing cancer, e.g. breast cancer -
XX Claim 2; Page 41-46; 49pp; English.

XX The invention provides an isolated antigen-presenting cell, which
XX expresses at least an immunogenic portion of a polypeptide that produces
XX an immune response to HER-2/neu protein. The antigen-presenting cells are
XX useful as vaccines for eliciting or enhancing an immune response to
XX HER-2/neu protein, particularly in treating or preventing malignancies in
XX which the HER-2/neu oncogene is associated. Specifically, these are
XX useful for treating or preventing cancer, e.g. breast cancer, ovarian,
XX colon, lung or prostate cancers. The present sequence represents
XX the human HER-2/neu protein (also known as p185 or c-erbB2).
XX Sequence 1255 AA;

Query Match 98.7%; Score 6714; DB 22; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLALLPPGAASCTGCTDMKRLRPASPTHLDMLRHLHYQGCVVQGNL 60
Db 1 MELAALCRWGLLALLPPGAASCTGCTDMKRLRPASPTHLDMLRHLHYQGCVVQGNL 60
QY 61 ELTYLPTNASLFLQDIQEVQVYLIHNAHQVPLQRLIRVRGTQLFEDNYALAVLDNG 1200

Db 61 ELTYLPTNASLFLQDIQEVQVYLIHNAHQVPLQRLIRVRGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLTORNPOLCYQDITLWKDILPHKKNOLA 180
Db 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLTORNPOLCYQDITLWKDILPHKKNOLA 180
QY 181 LTLIDNRRACHPCSPCKGRCWGESSEDCQSLTRTVCAAGCARCKGKPLPTDCCHEQC 240
Db 181 LTLIDNRRACHPCSPCKGRCWGESSEDCQSLTRTVCAAGCARCKGKPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVQYIKANSKFITGITERLYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVQYIKANSKFITGITERLYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHNEVTAEDGTORCEKCKPCARVCYGLGMEHLREVRATYSAN 360
Db 301 YNYLSTDVGSCTLVCPHNEVTAEDGTORCEKCKPCARVCYGLGMEHLREVRATYSAN 360
QY 361 IOEFAGCKKI FGSIAFLPESFDGDPASNTAPIQEOLOVFETLEETGYLYISAMPDLSL 420
Db 361 IOEFAGCKKI FGSIAFLPESFDGDPASNTAPIQEOLOVFETLEETGYLYISAMPDLSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLRELGLALIHHTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLRELGLALIHHTHLCFVHTV 480
QY 481 PWDQLFRNPHOALLHTANRPEDECVEGLACHQLCARGHCWGPGPTQCVCNCSQFLRGEC 540
Db 481 PWDQLFRNPHOALLHTANRPEDECVEGLACHQLCARGHCWGPGPTQCVCNCSQFLRGEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHBPCQPNQSVTCFGEADQCVACAHYKDPFPCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHBPCQPNQSVTCFGEADQCVACAHYKDPFPCVARC 600
QY 601 PSGVKPDLSPYMPKPFDEEGACQPCPINCTHSCVDLDDKGCPCAEQASPLTSIVSAVVG 660
Db 601 PSGVKPDLSPYMPKPFDEEGACQPCPINCTHSCVDLDDKGCPCAEQASPLTSIVSAVVG 660
QY 661 ILLVVLGVVFGILIKRROOKIKYTMRLLOTELVEPLTPSGAMPNQAOMRLKETEL 720
Db 661 ILLVVLGVVFGILIKRROOKIKYTMRLLOTELVEPLTPSGAMPNQAOMRLKETEL 720
QY 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780
QY 781 YVSRLLGICLTSTVQLVTQMPYGLLDHVRNRLGSDLLNWCMIAGKMSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTQMPYGLLDHVRNRLGSDLLNWCMIAGKMSYLEDVR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNWALSILRRRT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNWALSILRRRT 900
QY 901 HOSDWSYGVTVWELMTFCAKPYDGIPIAREIPDLLEKGERLPQPPICITDVMYMWKCM 960
Db 901 HOSDWSYGVTVWELMTFCAKPYDGIPIAREIPDLLEKGERLPQPPICITDVMYMWKCM 960
QY 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGSPASPLDSTFYKSLLEDMDGLVDA 1020
Db 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGSPASPLDSTFYKSLLEDMDGLVDA 1020
QY 1021 BEYLVPQOQFFCDPAPAGAGWVHRRSSSTRSGGDLTLGLEPSEERAPSPAPSEG 1080
Db 1021 BEYLVPQOQFFCDPAPAGAGWVHRRSSSTRSGGDLTLGLEPSEERAPSPAPSEG 1080
QY 1081 AGSDVFDGDLGMAKGLQSLPTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAKGLQSLPTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
QY 1141 NQPDVRQPSPREGPLPAARPAATLERAKTILSPGKGVVVDVAFGGAVENPEYLTPO 1200

Db 1141 NQDVRPQPPSPREGPLPAARPAATLRLPKTLSPKNGVVDVFAFGGAVENPEYLTPQ 1200
 QY 1201 GGAAPQHPHPPAPSPADNLYYWDQDPPRGAPSPSTFKGTPTAENPEYLGDDVPV 1255
 Db 1201 GGAAPQHPHPPAPSPADNLYYWDQDPPRGAPSPSTFKGTPTAENPEYLGDDVPV 1255
 RESULT 10
 AAG88267
 ID AAG88267 standard; Protein; 1255 AA.
 XX AAG88267;
 AC AAG88267;
 DT 11-SEP-2001 (first entry)
 XX
 DE HER2/neu amino acid sequence.
 XX Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
 KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
 KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
 XX Homo sapiens.
 OS
 XX WC200141787-A1.
 PN 14-JUN-2001.
 PD
 XX 11-DEC-2000; 2000WO-US33591.
 XX 10-DEC-1999; 99US-0458299.
 XX (EPIM-) EPIMUNE INC.
 PA
 XX Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
 PI Keogh E;
 XX WPI; 2001-374995/39.
 DR
 XX An isolated prepared HER2/neu epitope useful in a vaccine for inducing
 PT cellular immune responses for the prevention and treatment of cancer -
 XX Disclosure; Page 15; 199pp; English.
 PS
 XX The present invention describes isolated prepared HER2/neu epitopes (I).
 CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
 CC culture in vitro and binds to a complex of an epitope (I), bound to a
 CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
 CC and a second epitope and the peptide is less than 50 contiguous amino
 CC acids that have 100% identity with a native peptide sequence of HER2/neu;
 CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
 CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
 CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
 CC and immunostimulant activities, and can be used in vaccines. (I), (II)
 CC and (III) are useful for inducing cellular immune responses for the
 CC prevention and treatment of cancer. (I) and (II) are useful for
 CC monitoring or evaluating an immune response to a tumour-associated
 CC antigen when incubated with a T lymphocyte sample form a patient and
 CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
 CC based vaccines mean that immunosuppressive epitopes that may be present
 CC in whole antigens may be avoided. Selected epitopes may be combined to
 CC enhance immunogenicity. The possible pathological side effects caused by
 CC infectious agents or whole protein antigen is eliminated. The vaccine
 CC provides the ability to direct and focus an immune response to multiple
 CC selected antigens from the same pathogen. Epitope-based anti-tumour
 CC vaccines provides the opportunity to combine epitopes derived from
 CC multiple tumour-associated molecules addressing the problem of tumour-
 CC tumour variability and reducing the likelihood of tumour escape due to
 CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
 CC the exemplification of the present invention.
 XX Sequence 1255 AA;

Query Match 98.7%; Score 6714; DB 22; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;
 Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;
 QY 1 MELAALCRWGLLALLPPGAASCTGCTDMKRLRPASPETHLDMRLHLYQGCVVQGNL 60
 Db 1 MELAALCRWGLLALLPPGAASCTGCTDMKRLRPASPETHLDMRLHLYQGCVVQGNL 60
 QY 61 ELTYLPTNASLFLQDIQEVQGVYLIHQNQVQVPLQRLIRVGTQLFDNTALAVLDNG 120
 Db 61 ELTYLPTNASLFLQDIQEVQGVYLIHQNQVQVPLQRLIRVGTQLFDNTALAVLDNG 120
 QY 121 DPLNNTTPTVGTASPGGLRELQRLSLEILKGGVLTORNPOLCYQDITLAKDIFHKNNOLA 180
 Db 121 DPLNNTTPTVGTASPGGLRELQRLSLEILKGGVLTORNPOLCYQDITLAKDIFHKNNOLA 180
 QY 181 LTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSILTRITVCAGGCARCKGKPLPTDCCHEQC 240
 Db 181 LTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSILTRITVCAGGCARCKGKPLPTDCCHEQC 240
 QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVQYIKANSKFIGITELRYTFGASCVTACP 300
 Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVQYIKANSKFIGITELRYTFGASCVTACP 300
 QY 301 YNYLSTDVGSCTLVCPHNOEVTAEQTCRCKSPCARVCYGLGMEHLREVRAVTSAN 360
 Db 301 YNYLSTDVGSCTLVCPHNOEVTAEQTCRCKSPCARVCYGLGMEHLREVRAVTSAN 360
 QY 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFEETLEETIGLYLISAWPDSL 420
 Db 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFEETLEETIGLYLISAWPDSL 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSLTIQGLGISWGLRSLRELGSGLALIHNNHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSLTIQGLGISWGLRSLRELGSGLALIHNNHLCFVHTV 480
 QY 481 PWDQLFRPHQALHTANRPEDECYVGEGLACHOLCARGHCWGPCTQVCNCSOFLRGQBC 540
 Db 481 PWDQLFRPHQALHTANRPEDECYVGEGLACHOLCARGHCWGPCTQVCNCSOFLRGQBC 540
 QY 541 VVECRVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFPFCVARC 600
 Db 541 VVECRVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFPFCVARC 600
 QY 601 PSGVKPDLISYMPWPKFPEDEGACQPCINCTHSCVDLDDKGPAPQASPLTSIISAVVG 660
 Db 601 PSGVKPDLISYMPWPKFPEDEGACQPCINCTHSCVDLDDKGPAPQASPLTSIISAVVG 660
 QY 661 ILLVVVLGVVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
 Db 661 ILLVVVLGVVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
 QY 721 RKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGS 780
 Db 721 RKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGS 780
 QY 781 YVSRLLGLICTSTVQLVTQLMPEYGLLDHVRNRLGSLQDILLNMCQIAKMSYLEDVYR 840
 Db 781 YVSRLLGLICTSTVQLVTQLMPEYGLLDHVRNRLGSLQDILLNMCQIAKMSYLEDVYR 840
 QY 841 LVHRDLAARNVLKSPNHNKVTDFGLARLLDIDETEHADGKVPKKNMALSIILRRRT 900
 Db 841 LVHRDLAARNVLKSPNHNKVTDFGLARLLDIDETEHADGKVPKKNMALSIILRRRT 900
 QY 901 HOSDVWSVGVTVWEIMLTGAKPYDGI PAREIPDLLEKGERLPOPPICITDVYMWKCMW 960
 Db 901 HOSDVWSVGVTVWEIMLTGAKPYDGI PAREIPDLLEKGERLPOPPICITDVYMWKCMW 960
 QY 961 IDSECRPRFRELVSFSSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
 Db 961 IDSECRPRFRELVSFSSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
 QY 1021 EBYLVPOQGFCCPDPAAGAGMVHHRSSSTRSGGDLTLGLEPSEEPAPSLAPSE 1080

Db 1021 EYLVPQQFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
 Qy 1081 AGSDVFDGDLGMAKGLQSLPTHDPSPLOYSSEDPVPLPSETDGVVAPLTCSPQPEYV 1140
 Db 1081 AGSDVFDGDLGMAKGLQSLPTHDPSPLOYSSEDPVPLPSETDGVVAPLTCSPQPEYV 1140
 Qy 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKOVFAFGGAVENPEYLTQP 1200
 Db 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERPKTLSPGKNGVVKOVFAFGGAVENPEYLTQP 1200
 Qy 1201 GGAAPQHPPPAFPAFDNLYYWDQDPPERGAPPSTFKGPTAENPYLGLDVPV 1255
 Db 1201 GGAAPQHPPPAFPAFDNLYYWDQDPPERGAPPSTFKGPTAENPYLGLDVPV 1255

RESULT 11

AAE24067
 ID AAE24067 standard; Protein; 1255 AA.

AC AAE24067;

XX 23-SEP-2002 (first entry)

DT Human Her-2 protein.

XX Human; Her-2; epidermal growth factor receptor 2; infection; cancer;

KW hyperproliferative disorder; prophylaxis; inflammation; antisense;

KW tumour; gene therapy; phosphorothioate backbone.

XX Homo sapiens.

OS WO200222636-A1.

XX 21-MAR-2002.

XX 12-SEP-2001; 2001WO-US28572.

XX 15-SEP-2000; 2000US-0663834.

XX (ISIS-) ISIS PHARM INC.

XX Bennett CF, Cowser LM;

XX WPI; 2002-471192/50.

XX N-PSDB; AAD38904.

XX Novel antisense oligonucleotide which modulates the expression of Human
 PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
 PT inflammation or to prevent infection in humans -
 XX Example 13; Page 95-107; 116pp; English.

CC The invention relates to antisense compounds targetted to a nucleic
 CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
 CC that specifically hybridises with and inhibits the expression of Her2.
 CC Antisense compounds of the invention are used for treating diseases or
 CC conditions associated with Her2 such as hyperproliferative disorders
 CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
 CC neural or cardiac cancer. They are also useful prophylactically e.g.
 CC to prevent or delay infection, inflammation and tumour formation. The
 CC invention is also used in gene therapy. The present sequence is human
 CC Her-2 protein.

XX Sequence 1255 AA;

Query Match 98.7%; Score 6714; DB 23; Length 1255;
 Best Local Similarity 98.8%; Pred. No. 0;
 Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLPPGAASQVCTGDMKURLPASPEHLDMLRHLYQGQCVVQGNL 60

Db 1 MELAALCRWGLLLALLPPGAASQVCTGDMKURLPASPEHLDMLRHLYQGQCVVQGNL 60

Qy 61 ELYTLPTNASLSFLQDIEQVGYVLIHAHQVQVPLQRLRIVRGTQOLFEDNYALAVLDNG 120
 Db 61 ELYTLPTNASLSFLQDIEQVGYVLIHAHQVQVPLQRLRIVRGTQOLFEDNYALAVLDNG 120
 Qy 121 DPLNNTTPTVGTASPGGLREQLASLTIELKGGVLIQORNPOLCYQDTILWKDIHFHKNOLA 180
 Db 121 DPLNNTTPTVGTASPGGLREQLASLTIELKGGVLIQORNPOLCYQDTILWKDIHFHKNOLA 180
 Qy 181 LTLIDNTRSRACHPCSPCKSGRCWGESSEDCOSLRTVTCAGGCARCKGKPLPDDCCHEQC 240
 Db 181 LTLIDNTRSRACHPCSPCKSGRCWGESSEDCOSLRTVTCAGGCARCKGKPLPDDCCHEQC 240
 Qy 241 AAGCTGPKHSDDCLACLFHFNHSGICEHFALVYIKANSKFIGITELRYTFGASCVTACP 300
 Db 241 AAGCTGPKHSDDCLACLFHFNHSGICEHFALVYIKANSKFIGITELRYTFGASCVTACP 300
 Qy 301 YNYLSTDVGSCTLVCPPLHNOEVTAEQTCCKSKPCARVCYGLGMEHLREVRVTSAN 360
 Db 301 YNYLSTDVGSCTLVCPPLHNOEVTAEQTCCKSKPCARVCYGLGMEHLREVRVTSAN 360
 Qy 361 IQEFAGCKITFGSLAFLPESFDGDPASNTAPLOEQLOVPFETLEEITGYLYISAMPDLSL 420
 Db 361 IQEFAGCKITFGSLAFLPESFDGDPASNTAPLOEQLOVPFETLEEITGYLYISAMPDLSL 420
 Qy 421 DLSVFQNLQVIRGRILHNGAYSITLQGLGTSWLGSLRELGSGLALIHNTHLFCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSITLQGLGTSWLGSLRELGSGLALIHNTHLFCFVHTV 480
 Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAACHOLCARGHCWGPGPTQCVNCSOFLRGQEC 540
 Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAACHOLCARGHCWGPGPTQCVNCSOFLRGQEC 540
 Qy 541 VEECRVLQGLPREYVNRHCLPCHPECQONGSVTCFGEADOCVCAHYKDPFFCVARC 600
 Db 541 VEECRVLQGLPREYVNRHCLPCHPECQONGSVTCFGEADOCVCAHYKDPFFCVARC 600
 Qy 601 PSGVKPDLSPYIWKFPDEEGACQPCINCTHSCVDLDDKGCPAEQASPLTSIVSAVVG 660
 Db 601 PSGVKPDLSPYIWKFPDEEGACQPCINCTHSCVDLDDKGCPAEQASPLTSIVSAVVG 660
 Qy 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAMRILKETEL 720
 Db 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAMRILKETEL 720
 Qy 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLENTSPKANKIILDEAYVMAGVSP 780
 Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLENTSPKANKIILDEAYVMAGVSP 780
 Qy 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRLGSGQDLNWCQIAKMSYLEDVR 840
 Db 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRLGSGQDLNWCQIAKMSYLEDVR 840
 Qy 841 LVHRDLAARNVLKSPNNHVKITDFGLARLLDIDETEHADGGKVPIKWMALSIILRRRFT 900
 Db 841 LVHRDLAARNVLKSPNNHVKITDFGLARLLDIDETEHADGGKVPIKWMALSIILRRRFT 900
 Qy 901 HQSDVMSYGVTVWELMTFCAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMVMKCMW 960
 Db 901 HQSDVMSYGVTVWELMTFCAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMVMKCMW 960
 Qy 961 IDSECRPRFRELVSFSESRMARDPQRFVJQNEIDLGPASPLDSTFYRSLLEDDDDMDGLVDA 1020
 Db 961 IDSECRPRFRELVSFSESRMARDPQRFVJQNEIDLGPASPLDSTFYRSLLEDDDDMDGLVDA 1020
 Qy 1021 EYLVPOQGFCCDPDPAGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
 Db 1021 EYLVPOQGFCCDPDPAGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
 Qy 1081 AGSDVFDGDLGMAKGLQSLPTHDPSPLOYSSEDPVPLPSETDGVVAPLTCSPQPEYV 1140
 Db 1081 AGSDVFDGDLGMAKGLQSLPTHDPSPLOYSSEDPVPLPSETDGVVAPLTCSPQPEYV 1140
 Qy 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKOVFAFGGAVENPEYLTQP 1200

Db 1141 NQDVRQPPSPREGPLFAARPAAGATLERPKTSLSPGKNGVVKDVFAPGGAVENPEYLTPQ 1200
QY 1201 GGAAPDHPPPAFSPAFDNLYYWDQPPPERGAPSPSTFKGTPTAENPEYGLDVPV 1255
Db 1201 GGAAPDHPPPAFSPAFDNLYYWDQPPPERGAPSPSTFKGTPTAENPEYGLDVPV 1255

RESULT 12
AAE20479
ID AAE20479 standard; Protein; 1255 AA.
XX AAE20479;
AC AAE20479;
XX 01-JUL-2002 (first entry)
XX Human Her-2/neu protein.
XX Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
KW human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX OS Homo sapiens.

XX Key Location/Qualifiers
FT Region 1021..1030
FT /note= "Naturally processed HLA-B44-restricted epitope"

XX WO200214503-A2.
XX 21-FEB-2002.
XX 14-AUG-2001; 2001WO-US41733.
XX 14-AUG-2000; 2000US-225152P.
XX 28-SEP-2000; 2000US-236428P.
XX 21-FEB-2001; 2001US-270520P.
XX (CORI-) CORIXA CORP.
XX Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
PI McNeill PD, Vedvick TS;
XX WPI; 2002-280758/32.
DR N-PSDB; AAD32743.
XX Novel isolated Her-2/Neu polypeptide composition useful for therapy,
PT prevention and diagnosis of cancer, preferably breast cancer
XX Disclosure; Page 114-117; 129pp; English.

XX The invention relates to an isolated Her-2/Neu polypeptide composition
CC effective for eliciting an immune response. The invention is useful for
CC eliciting an immune response in a patient, where the patient is human
CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
CC The composition is useful for the therapy and diagnosis of cancer,
CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
CC and other compositions for the diagnosis, prevention and treatment of
CC human malignancies, for stimulating and/or expanding T cells specific for
CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a
CC patient. The invention is useful for stimulating a T cell response in a
CC human patient, as probe or primer for nucleic acid hybridisation, to
CC selectively form duplex molecules with complementary stretches of the
CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full
CC length gene from a suitable library, and to direct expression of a
CC polypeptide in appropriate host cells. The composition is useful in
CC prophylactic or therapeutic applications and for the treatment of cancer,
CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-
CC associated malignancies. The invention is useful in gene therapy. The
CC present sequence is human Her-2/neu protein.

XX Sequence 1255 AA;

Query Match 98.7%; Score 6714; DB 23; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKLRLPASPETHLDMLRHLYQGCVVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKLRLPASPETHLDMLRHLYQGCVVQGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQGVLIHAHQVQVPLQRLIRVRGTQLPEDNVALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQGVLIHAHQVQVPLQRLIRVRGTQLPEDNVALAVLDNG 120
QY 121 DPLNNTTPVTGASPGGLRELQRLSILTEILKGGVILQORNPOLCYODTILMKDIFHKKNOLA 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSILTEILKGGVILQORNPOLCYODTILMKDIFHKKNOLA 180
QY 181 LTLIDTNRSEACHPCSPCKGSRGCESEDQSLTRTVACGACGACRCKGKPLPTDCCHEOC 240
Db 181 LTLIDTNRSEACHPCSPCKGSRGCESEDQSLTRTVACGACGACRCKGKPLPTDCCHEOC 240
QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVYOYIKANSKFITELRYTTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVYOYIKANSKFITELRYTTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPPLHNOEVTAEQTCRCKSPCARVCYGLGMEHLREVRATVSAN 360
Db 301 YNYLSTDVGSCTLVCPPLHNOEVTAEQTCRCKSPCARVCYGLGMEHLREVRATVSAN 360
QY 361 IOEPFAGCKKI FGSFLAPLPESFDGDPASNTAPLOPELOQVFETLEETITGLYLTISAWPDSL 420
Db 361 IOEPFAGCKKI FGSFLAPLPESFDGDPASNTAPLOPELOQVFETLEETITGLYLTISAWPDSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRELGLSLALIHNNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRELGLSLALIHNNTHLCFVHTV 480
QY 481 PWDQLFRNPQALHTANRPEDECVGEGLAACHOLCARGHCWGPPTQCVNCSOFLRGQEC 540
Db 481 PWDQLFRNPQALHTANRPEDECVGEGLAACHOLCARGHCWGPPTQCVNCSOFLRGQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKOPPPFCVAC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKOPPPFCVAC 600
QY 601 PSGVKPDLSPYMPKPPDEEGACQPCPCINCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660
Db 601 PSGVKPDLSPYMPKPPDEEGACQPCPCINCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660
QY 661 ILLVVVLGVVFGILIKRROQKIRKVTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRROQKIRKVTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 721 RKVKVLGSGAFGVYKGIWIPGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGVYKGIWIPGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGP 780
QY 781 YVSRLLGICLTSTVOLVTOLMEYGLLDHVRNRRGLSGQDILLNWCQIAKMSYLEDVR 840
Db 781 YVSRLLGICLTSTVOLVTOLMEYGLLDHVRNRRGLSGQDILLNWCQIAKMSYLEDVR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKKNWMALESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKKNWMALESILRRRFT 900
QY 901 HQSDVMSYGVTVWELMTFCAKEYDGI PAREIPDLLEKGERLPQPPICTIDVYMIWKVM 960
Db 901 HQSDVMSYGVTVWELMTFCAKEYDGI PAREIPDLLEKGERLPQPPICTIDVYMIWKVM 960
QY 961 IDSECRPRELVSEFSSRMARDPQRFVITQNEDLGPASPLDSTFYRSLLEDDMGDLVDA 1020
Db 961 IDSECRPRELVSEFSSRMARDPQRFVITQNEDLGPASPLDSTFYRSLLEDDMGDLVDA 1020
QY 1021 EBYLVPOQGFCCPDPAAGAGMVHRRSSSTRSGGDLTLGLPSEEPAPSLAPSG 1080

Db 1021 EYLVPQQGFCPPAPAGAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGLGMAKGLSLPHDPSPLQYSEDPTVPLPSSTDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGLGMAKGLSLPHDPSPLQYSEDPTVPLPSSTDGVVAPLTCSPQPEYV 1140
Qy 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYLTPO 1200
Db 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPSPSTFGKPTTAENPEYLGLDVVP 1255
Db 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPSPSTFGKPTTAENPEYLGLDVVP 1255

RESULT 13

AAM51143
ID AAM51143 standard; Protein; 1255 AA.

AC AAM51143;
XX 17-JUN-2002 (first entry)
DE Human Her-2/neu oncogene-encoded p185 glycoprotein.
XX Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
KW tyrosine kinase; receptor; c-erbB2; gene therapy.
XX Homo sapiens.

Key Location/Qualifiers
FH Domain 1..653
FT /note= "extracellular domain"
FT Domain 676..1255
FT /note= "intracellular domain"
FT Domain 990..1255
FT /note= "phosphorylation domain"

XX W0200212341-A2.

XX 14-FEB-2002.

XX 03-AUG-2001; 2001WO-US24283.

XX 03-AUG-2000; 2000US-0632507.

XX (CORI-) CORIXA CORP.
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Cheever MA, Gheysen D;

XX WPI; 2002-241743/29.

XX N-PSDB; ABA92250.

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting
PT or enhancing an immune response to the protein, has Her-2/neu
PT extracellular domain fused to Her-2/neu intracellular or
PT phosphorylation domain

XX Claim 68; Fig 7; 141pp; English.

XX The present sequence is that of human Her-2/neu (p185 glycoprotein
CC or c-erbB2), an oncogenic self-protein and target for anti-cancer
CC vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
CC in a variety of cancers, including breast, ovarian, colon, lung and
CC prostate cancer. Her-2/neu is a member of the tyrosine kinase
CC family of receptor-like glycoproteins. It comprises an extracellular
CC domain with homology to the epidermal growth factor receptor
CC (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
CC intracellular domain that also shows homology to EGFR. Its
CC overexpression correlates with a poor prognosis in breast and
CC ovarian cancers. The invention provides Her-2/neu fusion
CC proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In
CC preferred fusion proteins, the extracellular domain of a Her-2/neu
CC protein is fused to a Her-2/neu intracellular domain or
CC phosphorylation domain (or its Deltapp fragment). An immune
CC response to Her-2/neu protein is elicited or enhanced by
CC administering the fusion protein in the form of a vaccine, or by
CC transfecting cells of an animal ex vivo with a nucleic acid
CC encoding the fusion protein, and delivering the transfected cells
CC to the animal. The fusion proteins, nucleic acids, and isolated
CC specific T-cells are useful for inhibiting the development of a
CC cancer, especially breast, ovarian, colon, lung or prostate cancer
CC in a patient. T cells that specifically react with a Her-2/neu
CC fusion protein can be used to remove tumour cells from a sample in
CC order to inhibit the development of cancer in a patient.

XX Sequence 1255 AA;

Query Match 98.7%; Score 6714; DB 23; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLPPGAASSTVCTGTDMLRLPASPTHLMRLHYQGCQVVGNL 60
Db 1 MELAALCRWGLLLALLPPGAASSTVCTGTDMLRLPASPTHLMRLHYQGCQVVGNL 60
Qy 61 ELYLPTNASLSFLQDIQEVQVLIHNVQVPLQRLRIVRGTLQFEDNVALAVLDNG 120
Db 61 ELYLPTNASLSFLQDIQEVQVLIHNVQVPLQRLRIVRGTLQFEDNVALAVLDNG 120
Qy 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKKNQLA 180
Db 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKKNQLA 180
Qy 181 LTLIDNTRSRACHPCSPMCKSGRCWGESSEDCQSLTRTYCAGGCARCKGPLPDDCCEQC 240
Db 181 LTLIDNTRSRACHPCSPMCKSGRCWGESSEDCQSLTRTYCAGGCARCKGPLPDDCCEQC 240
Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVQYIKANSKFGITELRTYFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVQYIKANSKFGITELRTYFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IQEFAGCKKIFGSLAPLPESFDGDPASNTAPLQPEQLQVFETLEEITGVLYISAWPDSLP 420
Db 361 IQEFAGCKKIFGSLAPLPESFDGDPASNTAPLQPEQLQVFETLEEITGVLYISAWPDSLP 420
Qy 421 DLSVFONLQVIRGRILHNGAYSLLTQGLGISWGLRSLRELGLALIHNNTHLCFVHTV 480
Db 421 DLSVFONLQVIRGRILHNGAYSLLTQGLGISWGLRSLRELGLALIHNNTHLCFVHTV 480
Qy 481 PMDQLFRNPHQALLHTANRPEDCEVGEGLACHQLCARGHCWGPGPTQCVCNCSQFLRGQEC 540
Db 481 PMDQLFRNPHQALLHTANRPEDCEVGEGLACHQLCARGHCWGPGPTQCVCNCSQFLRGQEC 540
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVCAHYKDPFPCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVCAHYKDPFPCVARC 600
Qy 601 PSGVKPDLVMPYIWKFPDDEGACOPCPINCTHSCVDLDDKGPAPORASPLTSIVSAVVG 660
Db 601 PSGVKPDLVMPYIWKFPDDEGACOPCPINCTHSCVDLDDKGPAPORASPLTSIVSAVVG 660
Qy 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
Qy 721 RKVKVLGSGAFGTVYKGIWIPGENVKIPVAIKVLRNTSPKANKIILDEAYMAGVGP 780
Db 721 RKVKVLGSGAFGTVYKGIWIPGENVKIPVAIKVLRNTSPKANKIILDEAYMAGVGP 780

QY 781 YVSELLGICLTSTVQLVTLMPYCLLDHVRNRRGRGLSGDOLLNWCQIAKMSYLEDDR 840
Db 781 YVSELLGICLTSTVQLVTLMPYCLLDHVRNRRGRGLSGDOLLNWCQIAKMSYLEDDR 840
QY 841 LVHRDLAARNVLKSPNHVKTDFGLARLLDIDETEHADGGKVPKIMMALESTLRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKTDFGLARLLDIDETEHADGGKVPKIMMALESTLRRFT 900
QY 901 HQSDVWSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICITDVMIMVKCM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICITDVMIMVKCM 960
QY 961 IDSECRPRFELVSEFSEMRDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMGDLVDA 1020
Db 961 IDSECRPRFELVSEFSEMRDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMGDLVDA 1020
QY 1021 EYLVPQOQFFCDDPAPACAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
Db 1021 EYLVPQOQFFCDDPAPACAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
QY 1081 AGSDVDFDGLGMAAGLQSLPHTDPSLQRYSEDPTVPLPSETDGYVAPLTCSPOPEYV 1140
Db 1081 AGSDVDFDGLGMAAGLQSLPHTDPSLQRYSEDPTVPLPSETDGYVAPLTCSPOPEYV 1140
QY 1141 NOPDVRPQPPSPREGPLPAARPAATLERAKTILSPGKNGVVKDVFAPGGAVENPEYLTPO 1200
Db 1141 NOPDVRPQPPSPREGPLPAARPAATLERAKTILSPGKNGVVKDVFAPGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPPSTFKGPTAENPEYLGLDVVPV 1255
Db 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPPSTFKGPTAENPEYLGLDVVPV 1255

RESULT 14

AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX AAU77114;
AC AAU77114;
XX 05-JUN-2002 (first entry)
DT Human Her-2/neu polypeptide.
XX Human; Her-2/neu; cytostatic; haematological malignancy; CML;
XX acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
XX chronic lymphocytic leukaemia; mveloma; non-Hodgkin's lymphoma; MDS;
XX Hodgkin's lymphoma; T cell therapy.
XX Homo sapiens.
OS WO200213847-A2.
XX 21-FEB-2002.
XX 13-AUG-2001; 2001WO-US25408.
XX 14-AUG-2000; 2000US-0638280.
XX 28-SEP-2000; 2000US-0675904.
XX (CORI-) CORIXA CORP.
XX Gaiger A, Cheever MA, Hand-zimmermann S;
PI WPI; 2002-280741/32.
XX N-PSDB; ABK10730.
XX Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
XX polypeptide
XX
PS Disclosure; Page 71-74; 74pp; English.
XX

CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;

Query Match 98.7%; Score 6714; DB 23; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLALLPPGAASCTGCTDMKLRLPASPETHLDMRLHYQGCQVQGNL 60
Db 1 MELAALCRWGLLALLPPGAASCTGCTDMKLRLPASPETHLDMRLHYQGCQVQGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQGVLIHQAQVPLQRLRIVRGTLQDFEDNALVALVDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQGVLIHQAQVPLQRLRIVRGTLQDFEDNALVALVDNG 120
QY 121 DPLNNTPTVTGASPGRLRLQSLTEILKGGVLIQORNQOLCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTPTVTGASPGRLRLQSLTEILKGGVLIQORNQOLCYQDTILWKDIFHKNNQLA 180
QY 181 LTLIDNRRACHPCSPMCKSGRCWGESSEDCQSLTRTYCAGGCARCKGPLTDCCHEQC 240
Db 181 LTLIDNRRACHPCSPMCKSGRCWGESSEDCQSLTRTYCAGGCARCKGPLTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVQYIKANSKFTIGITELRTYFAGSCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVQYIKANSKFTIGITELRTYFAGSCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHNLQEVTAEDGTQRCBCKSPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHNLQEVTAEDGTQRCBCKSPCARVCYGLGMEHLREVRVTSAN 360
QY 361 IQEFAGCKKIFGSLAFLESFDDGDPASNTAPLOPQLOVTFETLEETIGLYISAWPDSL 420
Db 361 IQEFAGCKKIFGSLAFLESFDDGDPASNTAPLOPQLOVTFETLEETIGLYISAWPDSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSILTIQGLIGLWGLSLRSLRELGLAIHNNHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSILTIQGLIGLWGLSLRSLRELGLAIHNNHLCFVHTV 480
QY 481 PWDQLFRNPQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
Db 481 PWDQLFRNPQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPEQOPNGSVTCFGEADQCACAHYKDPFPCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPEQOPNGSVTCFGEADQCACAHYKDPFPCVARC 600
QY 601 PSGVKPDLSPYMPIMWKFPDEEGACQPCPINCTHSCVDLDKQCPAEORASPLTSIIVSAVVG 660
Db 601 PSGVKPDLSPYMPIMWKFPDEEGACQPCPINCTHSCVDLDKQCPAEORASPLTSIIVSAVVG 660
QY 661 ILLVVVLGVVFGILIKRQKQKIRKYMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRQKQKIRKYMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVRENTSPKANKEILDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVRENTSPKANKEILDEAYVMAGVGP 780
QY 781 YVSRLLGICLTSTVQLVTLMPYCLLDHVRNRRGRGLSGDOLLNWCQIAKMSYLEDDR 840
Db 781 YVSRLLGICLTSTVQLVTLMPYCLLDHVRNRRGRGLSGDOLLNWCQIAKMSYLEDDR 840
QY 841 LVHRDLAARNVLKSPNHVKTDFGLARLLDIDETEHADGGKVPKIMMALESTLRRFT 900

Db 841 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEVHADGGKVPKIMMALESILRRFT 900
Qy 901 HQSDVRSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVKWM 960
Db 901 HQSDVRSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVKWM 960
Qy 961 IDSECRPRFRELVSFMRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGLVDA 1020
Db 961 IDSECRPRFRELVSFMRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGLVDA 1020
Qy 1021 EYLVPQGFCCPDPAAGAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
Db 1021 EYLVPQGFCCPDPAAGAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAEPLTCSQPBEYV 1140
Db 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAEPLTCSQPBEYV 1140
Qy 1141 NQPDVRPQPPSPREGPLPAARPAATLERAKTILSPGKNGVVVDVFAFGGAVENPEYLTPO 1200
Db 1141 NQPDVRPQPPSPREGPLPAARPAATLERAKTILSPGKNGVVVDVFAFGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPPPAFPAFDNLYWDQDPPPERGAPPSTFKGTPTAENPEYLGIDVPV 1255
Db 1201 GGAAPQHPPPAFPAFDNLYWDQDPPPERGAPPSTFKGTPTAENPEYLGIDVPV 1255

RESULT 15

AAR39568
ID AAR39568 standard; Protein; 1433 AA.

XX AC AAR39568;

XX DT 07-FEB-1994 (first entry)

XX DE Sequence of c-erbB-2 tumour antigen.

XX KW Tumour antigen; c-erbB-2; glycoprotein.

XX OS Homo sapiens.

XX PN WO9316185-A.

XX PD 19-AUG-1993.

XX PF 05-FEB-1993; 93WO-US01055.

XX PR 06-FEB-1992; 92US-0831967.

XX PA (CETU) CETUS ONCOLOGY CORP.

XX PA (CREA-) CREATIVE BIOMOLECULES INC.

XX PI Houston LL, Huston JS, Oppermann H, Ring DB;

XX DR WPI: 1993-272889/34.

XX DR N-PSDB; AAQ46083.

XX PT New single chain Fv polypeptide binding to C-erbB-2 tumour

XX PT antigen - for imaging or treating breast or ovarian cancer etc.

XX PS Disclosure; pages 48-54; 87pp; English.

XX CC c-erbB-2 refers to a protein antigen expressed on the surface of

XX CC tumour cells. such as breast and ovarian tumour cells, which is an

XX CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric

XX CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents

XX CC the location of a stop codon in AAQ46083.

XX SQ Sequence 1433 AA;

Query Match

Best Local Similarity 98.1%; Score 6671; DB 14; Length 1433;

Pred. No. 0;

Matches 1233; Conservative 5; Mismatches 17; Indels 0; Gaps 0;
Qy 1 MELAALCRGLLALLAPPAGAASTQVCTGDMKLRLPASPETHLDMLRHLYQGCQVQGNL 60
Db 1 MELAALCRGLLALLAPPAGAASTQVCTGDMKLRLPASPETHLDMLRHLYQGCQVQGNL 60
Qy 61 EUTYLPNTASLFLODIOEVQGVLIHAHQVQVPLQRLRIVRGTOLFEDNYALAVLDNG 120
Db 61 EUTYLPNTASLFLODIOEVQGVLIHAHQVQVPLQRLRIVRGTOLFEDNYALAVLDNG 120
Qy 121 DPLNNTPTVGTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTRILKWDIFHKKNOLA 180
Db 121 DPLNNTPTVGTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTRILKWDIFHKKNOLA 180
Qy 181 LTLIDNTRACHPCSPCKGSRGWESSEDCOSLTRTVCAGGCARCKGKPLPTDCCHQEC 240
Db 181 LTLIDNTRACHPCSPCKGSRGWESSEDCOSLTRTVCAGGCARCKGKPLPTDCCHQEC 240
Qy 241 AAGCTGPKHSDCLACILHFNHSGICELHCPALVQYIKANSKFICITELRTYFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACILHFNHSGICELHCPALVQYIKANSKFICITELRTYFGASCVTACP 300
Qy 301 YNYLSTDVSGSCTLVCPHMQEVTAEQTCRCKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVSGSCTLVCPHMQEVTAEQTCRCKSKPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IOEFAGCKTIFGSLAEPLPESFDGDPASNTAPLOPEHLQVFEETLEETGVLYISAWPDSL 420
Db 361 IOEFAGCKTIFGSLAEPLPESFDGDPASNTAPLOPEHLQVFEETLEETGVLYISAWPDSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLRSLRELGSGLALIHNNTHLSFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLRSLRELGSGLALIHNNTHLSFVHTV 480
Qy 481 PMDQLPFNPHQALLHTANRPEDECVGEGLAACHOLCARGHCWGPGTQCVNCSOFLRGQEC 540
Db 481 PMDQLPFNPHQALLHTANRPEDECVGEGLAACHOLCARGHCWGPGTQCVNCSOFLRGQEC 540
Qy 541 VVEECRVQLGPREYNNARHCLPCHPECPQNGSVTCFGEADQCVACAHKDPFPCVARC 600
Db 541 VVEECRVQLGPREYNNARHCLPCHPECPQNGSVTCFGEADQCVACAHKDPFPCVARC 600
Qy 601 PSQVKPDLGYMPIWKPFPDEBEGACQPCPINCTHSCVDLDDKGPAPQASPLTSISAVVG 660
Db 601 PSQVKPDLGYMPIWKPFPDEBEGACQPCPINCTHSCVDLDDKGPAPQASPLTSISAVVG 660
Qy 661 ILLVVVLGVVFGILIKRROQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRROQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Qy 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKILDEAYVMAGVSP 780
Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKILDEAYVMAGVSP 780
Qy 781 YVSRLLGICLTSTVQLVTLQMPYCLLDHVRENRGLSGQDLINWCWQIAKNGSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTLQMPYCLLDHVRENRGLSGQDLINWCWQIAKNGSYLEDVR 840
Qy 841 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEVHADGGKVPKIMMALESILRRFT 900
Db 841 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEVHADGGKVPKIMMALESILRRFT 900
Qy 901 HQSDVRSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVKWM 960
Db 901 HQSDVRSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVKWM 960
Qy 961 IDSECRPRFRELVSFMRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGLVDA 1020
Db 961 IDSECRPRFRELVSFMRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGLVDA 1020
Qy 1021 EYLVPQGFCCPDPAAGAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
Db 1021 EYLVPQGFCCPDPAAGAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080


```

Qy 1081 AGSDVFDGDLGMGAAGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMGAAGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO 1200
Db 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAFSPAFDNLYYWDQDPPPERGAPSTFKGTPTAENPEYLGLDVVPV 1255
Db 1201 GGAAPQHPHPPAFSPAFDNLYYWDQDPPPERGAPSTFKGTPTAENPEYLGLDVVPV 1255

```

Search completed: July 22, 2003, 08:41:01
Job time : 43.9774 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 37.9774 Seconds
(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-59-73-12

Perfect score: 6809

Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYIGLDVVPV 1255

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-embl/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-embl/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-embl/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-embl/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-embl/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-embl/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-embl/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-embl/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-embl/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-embl/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-embl/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-embl/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-embl/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-embl/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-embl/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-embl/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-embl/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-embl/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-embl/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-embl/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-embl/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-embl/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-embl/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6735	98.9	1255	21	AA92620 Human heregulin 2
2	6735	98.9	1255	22	AAE12130 Human tyrosine kin
3	6735	98.9	1255	22	AAB60167 HER2 transgene pla
4	6735	98.8	1255	23	AAU74545 Human HER2 (ErbB2)
5	6729	98.8	1255	17	AAW01111 HER-2/neu protein.
6	6729	98.8	1255	20	AAW92406 Human HER-2/neu on
7	6729	98.8	1255	21	AAAB21198 Human HER-2/neu pr
8	6729	98.8	1255	21	AAAY84780 Amino acid sequenc
9	6729	98.8	1255	22	AAB85458 Human HER-2/neu pr
10	6729	98.8	1255	22	AAG88267 HER2/neu amino aci

11	6729	98.8	1255	23	AAE24067 Human Her-2 protei
12	6729	98.8	1255	23	AAE20479 Human Her-2/neu pr
13	6729	98.8	1255	23	AAAM51143 Human Her-2/neu on
14	6729	98.8	1255	23	AAU77114 Human Her-2/neu po
15	6686	98.2	1433	14	AAAR39568 Sequence of c-erbB
16	6565	96.4	1223	23	AAU98923 Human breast cance
17	6412	94.2	1200	21	AAAB21208 Human HER-2/neu pr
18	5929.5	87.1	1256	21	AAAB21199 Rat HER-2/neu prot
19	5929.5	87.1	1256	23	AAAM51144 Rat Her-2/neu onco
20	5909.5	86.8	1256	21	AAAB21206 Mouse Her-2/neu pr
21	5909.5	86.8	1256	22	AAAG62860 Amino acid sequenc
22	5909.5	86.8	1256	23	AAAM51151 Mouse Her-2/neu on
23	4815	70.7	919	21	AAAB21203 Human HER-2/neu fu
24	4815	70.7	919	23	AAAM51148 Her-2/neu extracel
25	4059.5	59.6	920	23	AAAM51152 Mouse Her-2/neu ex
26	4059.5	59.6	926	23	AAAM51153 Mouse Her-2/neu ex
27	3699	54.3	712	21	AAAB21204 Human HER-2/neu fu
28	3699	54.3	712	23	AAAM51149 Her-2/neu extracel
29	3553	52.2	782	18	AAAW19764 Her2-GM-CSF immuno
30	3551	52.2	653	21	AAAB21200 Extracellular HER-
31	3551	52.2	653	23	AAAM51145 Human ErbB2 oncopr
32	3513	51.6	645	22	AAAB60408 Human ErbB2 extrac
33	3513	51.6	645	22	AAAB61593 DC86CFV-erbB2EC fu
34	3448	50.6	951	21	AAAY44993 Extracellular port
35	3345	49.1	624	11	AAAR08222 Amino acid sequenc
36	3117	45.8	1210	21	AAAB19259 Human EGF receptor
37	3117	45.8	1210	21	AAAY50616 Human Her-1 protei
38	3117	45.8	1210	23	AAAB23019 Human epidermal gr
39	3117	45.8	1210	23	AAAM50768 Amino acid sequenc
40	3115	45.7	1210	22	AAAB68420 Human protein for
41	3084	45.3	583	23	AAAE20481 Human protein for
42	3084	45.3	587	23	AAAE20481 Human protein for
43	3083	45.3	589	23	AAAE20481 Human protein for
44	3083	45.3	600	23	AAAE20482 Human protein for
45	3076	45.2	1210	23	ABP51768 Human epidermal gr

ALIGNMENTS

RESULT 1
AA92620
ID AA92620 standard; Protein; 1255 AA.
XX
AC AA92620;
XX
DT 10-AUG-2000 (first entry)
XX
DE Human heregulin 2 (Her2).
XX
KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
XX
OS Homo sapiens.
XX
PH Key
FT Domain
FT Location/Qualifiers
FT /label= N-terminal
FT /note= "mature polypeptide"
FT 5...25
FT Region
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 59...73
FT Region
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 103...117
FT Region
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 149...163
FT Region
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 174...323
FT Domain

FT Region /label= Cysteine_rich_domain
 FT 210..224
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 250..264
 FT Region /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 324..483
 FT /label= Ligand_binding_domain
 FT 325..339
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 369..383
 FT Region /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 465..479
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 484..623
 FT Domain /label= Cysteine_rich_domain
 FT 579..593
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 624..654
 FT Domain /label= Transmembrane_domain
 FT 632..652
 FT Region /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 653..667
 FT Region /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 655..1010
 FT Domain /label= Tyrosine_kinase_domain
 FT 661..675
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 695..709
 FT Region /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 710..730
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 1011..1235
 FT Domain /label= C-terminal_domain
 XX WO200020027-A2.
 EN 13-APR-2000.
 XX 05-OCT-1999; 99WO-DK00525.
 XX 05-OCT-1998; 98DK-0001261.
 PR 20-OCT-1998; 98US-0105011.
 XX (MEBI-) M & E BIOTECH AS.
 XX Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
 PI Gautam A, Birk P, Karlsson G;
 XX WPI; 2000-349917/30.
 DR N-PSDB; AAA09455.
 DR Inducing immune responses to weakly immunogenic, tumor associated
 PT peptide antigens for the treatment of breast and prostate cancer
 XX Claim 62; Page 193-198; 220pp; English.
 PS This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
 CC Her2 can be used in the claimed method as an autovaccine to induce a CTL
 CC response. Subdominant CTL epitopes, antibody binding regions and
 CC cysteine residues involved in disulfide bonds are preserved in the
 CC immunogenized forms. Regions suitable for the insertion of foreign T
 CC helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
 CC cell-associated peptide antigens (PA) such as those associated with
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
 CC The method comprises effecting simultaneous presentation by antigen
 CC producing cells (APCs) of the animals immune system of: (1) at least 1
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
 CC B-cell group derived from the cell-associated PA; and (2) at least 1
 CC first T helper cell group which is foreign to the animal. Analogues of
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
 CC part of all known and predicted CTL and B-cell epitopes of the respective
 CC PA and including at least one foreign T helper epitope are also claimed.
 CC The method is used to treat prostate, prostate/breast or breast cancer
 CC when the PA is human PSM, FGF8b and Her2, respectively.

XX Sequence 1255 AA;

Query Match 98.9%; Score 6735; DB 21; Length 1255;
 Best Local Similarity 99.0%; Pred. No. 0;
 Matches 1242; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLPASPETHLDMLRHLVGGCQVQGNL 60
 Db 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLPASPETHLDMLRHLVGGCQVQGNL 60

Qy 61 ELTYLPTNASLSFLQDIEVQVQYIKANSKFITGTELQRLRIVRGTLQFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLSFLQDIEVQVQYILAHNQVRQVPLQRLRIVRGTLQFEDNYALAVLDNG 120

Qy 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTTLWKIDIFHNQOLA 180
 Db 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTTLWKIDIFHNQOLA 180

Qy 181 LTLIDTNRSRACHPCSPMKGRCWGESSEDCQSLRTVCAGGCARCKGLPTDCCHEQC 240
 Db 181 LTLIDTNRSRACHPCSPMKGRCWGESSEDCQSLRTVCAGGCARCKGLPTDCCHEQC 240

Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 360
 Db 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQEQLOVFTELEEITGYLIYSAMPDPL 420
 Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQEQLOVFTELEEITGYLIYSAMPDPL 420

Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLGISWLGRLSRLRELGLSLIHNHNLFCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLGISWLGRLSRLRELGLSLIHNHNLFCFVHTV 480

Qy 481 PWDQLFRNPQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGPTQCVNCSQFLRQEC 540
 Db 481 PWDQLFRNPQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGPTQCVNCSQFLRQEC 540

Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFPGPADQCVACAHYKDPFCVARC 600
 Db 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFPGPADQCVACAHYKDPFCVARC 600

Qy 601 PSGVKPDLSYMPITWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSTVSAVVG 660
 Db 601 PSGVKPDLSYMPITWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSTVSAVVG 660

Qy 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRLLEQTELVEPLTPSGAMPNQAMRILKETEL 720
 Db 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRLLEQTELVEPLTPSGAMPNQAMRILKETEL 720

Qy 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETILDEAYVMAGVGP 780
 Db 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETILDEAYVMAGVGP 780

QY 781 YVRLGLCLSTVLTQMPYGCCLLDHVRNRLGSLQDLNWCMIKAGMSYLEDVR 840
 Db YVRLGLCLSTVLTQMPYGCCLLDHVRNRLGSLQDLNWCMIKAGMSYLEDVR 840
 QY 841 LVHRDLAARNVLKSPNHNKVTDFGLARLLDDIDETEXHADGKVPKIMMALESILRRFT 900
 Db LVHRDLAARNVLKSPNHNKVTDFGLARLLDDIDETEXHADGKVPKIMMALESILRRFT 900
 QY 901 HQSDWMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICTIDVYIMVKWM 960
 Db HQSDWMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICTIDVYIMVKWM 960
 QY 961 IDSECRPRFRLVSEFSMARDPQRFVVIQNEIDGPASPLSTFYRSLLDDMGDLVDA 1020
 Db IDSECRPRFRLVSEFSMARDPQRFVVIQNEIDGPASPLSTFYRSLLDDMGDLVDA 1020
 QY 1021 EYLVPQGGFFCPDPAFCAGGMVHRHRSSTRSGGDLTLGLPSEBEAPRSLAPSEG 1080
 Db EYLVPQGGFFCPDPAFCAGGMVHRHRSSTRSGGDLTLGLPSEBEAPRSLAPSEG 1080
 QY 1081 AGSDVFDGDLGWAAGKGLQSLTPHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 Db AGSDVFDGDLGWAAGKGLQSLTPHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 QY 1141 NOPDVRPQPPSPREGPLPAARPAATLRAKTLSPGKNGVVDVPAFCGAVENPEYLTPO 1200
 Db NOPDVRPQPPSPREGPLPAARPAATLRAKTLSPGKNGVVDVPAFCGAVENPEYLTPO 1200
 QY 1201 GGAAPQHPHPPAFSPADNLVYWDQDPPERCAPPSTPKGTPTAENPEYLGLDVVP 1255
 Db GGAAPQHPHPPAFSPADNLVYWDQDPPERCAPPSTPKGTPTAENPEYLGLDVVP 1255

RESULT 2

AAE12130
 ID AAE12130 standard; Protein; 1255 AA.

AC AAE12130;

DT 18-DEC-2001 (first entry)

DE Human tyrosine kinase-type receptor, HER-2.

KW Therapeutic compound; major histocompatibility complex; vaccine;
 KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
 KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
 KW antigen presenting cell; human; tyrosine kinase-type receptor.

OS Homo/sapiens.

XX Key Location/Qualifiers
 FH 774..782
 FT Region /note= "Antigenic epitope"

XX WO200168677-A2.

XX 20-SEP-2001.

XX 16-MAR-2001; 2001WO-US40328.

XX 16-MAR-2000; 2000US-0527487.

XX (GENZ) GENZYME CORP.

XX Nicolette CA;

XX WPI; 2001-616284/71.

DR N-PSDB; AAD19731.

XX Novel synthetic therapeutic compound for inducing immune response and
 PT for use in adoptive immunotherapy, has enhanced binding to major
 PT histocompatibility molecules and enhanced immunoregulatory properties

XX Claim 4; Page 63-67; 69pp; English.
 XX The invention relates to synthetic therapeutic compounds (antigenic
 CC peptides) with enhanced binding to major histocompatibility complex
 CC (MHC) molecules and enhanced immunoregulatory properties relative
 CC to their natural counterparts. Compounds of the invention are useful
 CC for inducing an immune response in a subject and for use in adoptive
 CC immunotherapy. They are useful as components of anti-cancer vaccines
 CC and to expand immune effector cells that are specific for cancers
 CC characterised by expression of the breast cancer antigen, HER-2.
 CC Polynucleotides that encode peptides of the invention are useful as
 CC hybridisation probes and as primers for the detection of genes of gene
 CC transcripts that are expressed in antigen presenting cells (APCs), to
 CC confirm transduction of polynucleotides into host cells. The present
 CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
 CC of the invention are designed based on the HER-2 antigenic peptide
 CC (774-782).

XX SQ Sequence 1255 AA;

Query Match 98.9%; Score 6735; DB 22; Length 1255;
 Best Local Similarity 99.0%; Pred. No. 0;
 Matches 1242; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALLPPGAASQVCTGTDMLRLPASPEHDLMLRHLVQGCQVQGNL 60

Db 1 MELAALCRWGLLALLPPGAASQVCTGTDMLRLPASPEHDLMLRHLVQGCQVQGNL 60

QY 61 ELYLPTNASLSFLQDIOEQVQYIKANSKFIGITELQRLIRVRGTQLFEDNYALAVLDNG 120

Db 61 ELYLPTNASLSFLQDIOEQVQYIKANSKFIGITELQRLIRVRGTQLFEDNYALAVLDNG 120

QY 121 DPLNNTTPVTGASPGGLRELQLESLEILKGGVLIORNPOLCYQDTILWKDIHKNQOLA 180

Db 121 DPLNNTTPVTGASPGGLRELQLESLEILKGGVLIORNPOLCYQDTILWKDIHKNQOLA 180

QY 181 LTLIDNRSRACHPCSPMCKGSRGWGSESSDCQSLTRTVCCAGCARCKGLPTDCCHEQC 240

Db 181 LTLIDNRSRACHPCSPMCKGSRGWGSESSDCQSLTRTVCCAGCARCKGLPTDCCHEQC 240

QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMNPPEGRYTFGASCVTACP 300

Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMNPPEGRYTFGASCVTACP 300

QY 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKPCARVCYGLGWEHLREVRVTSAN 360

Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKPCARVCYGLGWEHLREVRVTSAN 360

QY 361 IQEFAGCKKIFGSLAFPLPSFDGDPASNTAPLOEQLOVPFETLEEITGYLYISAWPDSL 420

Db 361 IQEFAGCKKIFGSLAFPLPSFDGDPASNTAPLOEQLOVPFETLEEITGYLYISAWPDSL 420

QY 421 DLSVFQNLQVIRGRIILHNGAYSILTLQGLISWLGSLRLSRLGSLALIHNTLHCFVHTV 480

Db 421 DLSVFQNLQVIRGRIILHNGAYSILTLQGLISWLGSLRLSRLGSLALIHNTLHCFVHTV 480

QY 481 PWDOLFERNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPGTQCVCNSOFLRGQEC 540

Db 481 PWDOLFERNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPGTQCVCNSOFLRGQEC 540

QY 541 VEECRVLQGLFREYVVARHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFPCVARC 600

Db 541 VEECRVLQGLFREYVVARHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFPCVARC 600

QY 601 PSGVKPDLSPYPIWKFPDEEGACQPCPINCTHSCVDLDDKGCFAEQASPLTSIVSAVVG 660

Db 601 PSGVKPDLSPYPIWKFPDEEGACQPCPINCTHSCVDLDDKGCFAEQASPLTSIVSAVVG 660

QY 661 ILLVVLGVVFGILIKRQOKIRKYTMRLLOSTELVEPLTPSGAMPNQAQMRILKETEL 720

Db 661 ILLVVLGVVFGILIKRQOKIRKYTMRLLOSTELVEPLTPSGAMPNQAQMRILKETEL 720

```

Qy 721 RKVKVLSGARGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKIILDEAYNVAGVSP 780
Db 721 RKVKVLSGARGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKIILDEAYNVAGVSP 780
Qy 781 YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRRGLSQDLLNWCQIAKGMYSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRRGLSQDLLNWCQIAKGMYSYLEDVR 840
Qy 841 LVHRDLAARNVVKSPNHVKITDGLARLLDIDETEHADGGKVPKIMWMALESILRRFT 900
Db 841 LVHRDLAARNVVKSPNHVKITDGLARLLDIDETEHADGGKVPKIMWMALESILRRFT 900
Qy 901 HQSDVWSYGVTVWELMTFGAPYDIPAREIPDLLEKGERLPQPPICITIDVYIMVWKWM 960
Db 901 HQSDVWSYGVTVWELMTFGAPYDIPAREIPDLLEKGERLPQPPICITIDVYIMVWKWM 960
Qy 961 IDSECRPRFRELVSFSESMARDPQRFVVIQNEIDLGASPLDSTFYRSILLEDDMGDLVDA 1020
Db 961 IDSECRPRFRELVSFSESMARDPQRFVVIQNEIDLGASPLDSTFYRSILLEDDMGDLVDA 1020
Qy 1021 BEYLVPOQGFPCPDPAAGAGWVHRRSSSTRSGGDLTLGLEPSEBEAPRSPAPSE 1080
Db 1021 BEYLVPOQGFPCPDPAAGAGWVHRRSSSTRSGGDLTLGLEPSEBEAPRSPAPSE 1080
Qy 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTQ 1200
Db 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTQ 1200
Qy 1201 GGAAPQHPHPPAFSPAFNLVYWDODPPERGAPSTFKGTPTAENPEYGLDVPV 1255
Db 1201 GGAAPQHPHPPAFSPAFNLVYWDODPPERGAPSTFKGTPTAENPEYGLDVPV 1255

```

RESULT 3

AAB60167 standard; Protein; 1255 AA.

AC AAB60167;

DT 03-APR-2001 (first entry)

XX HER2 transgene plasmid construct encoded protein.

XX Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer; antibody.

OS Homo sapiens.

OS Synthetic.

XX WO200100244-A2.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000WO-US17229.

XX 25-JUN-1999; 99US-0141316.

XX 16-MAR-2000; 2000US-0189844.

XX (GETH) GENENTECH INC.

XX Erickson S, Schwall R;

XX WPI; 2001-061962/07.

XX N-PSDB; AAF24297.

XX Treating tumors, particularly breast cancers, which overexpress an ErbB receptor and does not respond to an anti-ErbB antibody, comprises conjugating the antibody to a maytansinoid -

PS Example 3; Fig 4; 92pp; English.

XX The present invention provides a method of treating cancer by administering a conjugate of anti-ErbB antibody with a maytansinoid. In particular, the antibody is directed against ErbB2 (also known as HER2 and p185neu). The method is particularly useful in the treatment of breast, ovarian, stomach, endometrial, salivary gland, kidney, colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.

XX Sequence 1255 AA;

Query Match 98.9%; Score 6735; DB 22; Length 1255;
 Best Local Similarity 99.0%; Pred. No. 0;
 Matches 1242; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

```

Qy 1 MELAALCRWGLLALPPGAASQVCTGTMKRLPASPTHLDMLHLHYQGCVVQGNL 60
Db 1 MELAALCRWGLLALPPGAASQVCTGTMKRLPASPTHLDMLHLHYQGCVVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQYIKANSKFITELQRLRVGTQOLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQYILAHNOVRQVPLQRLRVGTQOLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDI FHKNNQLA 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDI FHKNNQLA 180
Qy 181 LTIIDNRSRACHPCSPMKSGRCWGESSEDCOSLTRTVCGAGCARCKGLPTDCHEQC 240
Db 181 LTIIDNRSRACHPCSPMKSGRCWGESSEDCOSLTRTVCGAGCARCKGLPTDCHEQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVSGCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVSGCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IQBFAGCKKIFGSLAFIPESFDGDPASNTAPLQEQLOVFETLEEITGYLIYISAWPSLP 420
Db 361 IQBFAGCKKIFGSLAFIPESFDGDPASNTAPLQEQLOVFETLEEITGYLIYISAWPSLP 420
Qy 421 DLSVFQNLQVIRGILHNGAYSILTQGLGISWLGRLSRLGSLALIHNTLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGILHNGAYSILTQGLGISWLGRLSRLGSLALIHNTLHLCFVHTV 480
Qy 481 PWDQLFNPHQALLHTANRPEDECYGEGLACHQLCARGHCWGPGTQCVNCSQFLRGOEC 540
Db 481 PWDQLFNPHQALLHTANRPEDECYGEGLACHQLCARGHCWGPGTQCVNCSQFLRGOEC 540
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFCVARC 600
Qy 601 PSGVKPDLSYMPIWKFPDEEGACOPCPINCTHSCVDLDDKGCPEAQRASPLTSIVSAVVG 660
Db 601 PSGVKPDLSYMPIWKFPDEEGACOPCPINCTHSCVDLDDKGCPEAQRASPLTSIVSAVVG 660
Qy 661 ILLVVLGVVFGIILIKRROQKIRKYTMRLLIQETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVVLGVVFGIILIKRROQKIRKYTMRLLIQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKIILDEAYNVAGVSP 780
Db 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKIILDEAYNVAGVSP 780
Qy 781 YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRRGLSQDLLNWCQIAKGMYSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRRGLSQDLLNWCQIAKGMYSYLEDVR 840
Qy 841 LVHRDLAARNVVKSPNHVKITDGLARLLDIDETEHADGGKVPKIMWMALESILRRFT 900

```

Db 841 LVHRDLARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPDKWMALESILRRPT 900
 Qy 901 HQSDVMSYGVTVWELMTGAKPYDIPAREIPDLLEKGERLPQPPICTIDVYIMVKWM 960
 Db 901 HQSDVMSYGVTVWELMTGAKPYDIPAREIPDLLEKGERLPQPPICTIDVYIMVKWM 960
 Qy 961 IDSECRPRFRELVSFMSMDPQRFVVIQNEIDLPASPLDSTFYRSLLDDMGDLVDA 1020
 Db 961 IDSECRPRFRELVSFMSMDPQRFVVIQNEIDLPASPLDSTFYRSLLDDMGDLVDA 1020
 Qy 1021 EYLVPQGGFFCPDPAAGAGGVHRRHSSSTRSGGGDLTLGLPSEBEAPRSPPLASEG 1080
 Db 1021 EYLVPQGGFFCPDPAAGAGGVHRRHSSSTRSGGGDLTLGLPSEBEAPRSPPLASEG 1080
 Qy 1081 AGSDVFDGDLGMAAGKGLQSLTPHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 Db 1081 AGSDVFDGDLGMAAGKGLQSLTPHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 Qy 1141 NOPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVDVPFAGGAVENPEYLTPO 1200
 Db 1141 NOPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVDVPFAGGAVENPEYLTPO 1200
 Qy 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPRGAPPSTFKGTPTAENPEYLGDDVPV 1255
 Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPRGAPPSTFKGTPTAENPEYLGDDVPV 1255

RESULT 4

AAU74545
 ID AAU74545 standard; Protein; 1255 AA.
 XX
 AC AAU74545;
 XX
 DT 23-APR-2002: (first entry)
 XX
 DE Human HER2 (ErbB2) polypeptide.
 XX
 KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;
 KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
 KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
 KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
 KW glial disorder; astrocytal disorder; hypothalamic disorder;
 KW glandular disorder; macrophagal disorder; epithelial disorder;
 KW stromal disorder; blastocoealic disorder; inflammatory disorder;
 KW angiogenic disorder; immunological disorder.

OS Homo sapiens.

PN US2002001587-A1.

XX 03-JAN-2002.

XX 16-MAR-2001; 2001US-0811123.

XX 16-MAR-2000; 2000US-189844P.

PR 05-OCT-2000; 2000US-238327P.

XX (ERIC/) ERICKSON S.

PA (SCHW/) SCHWALL R.

PA (SLIW/) SLIWKOWSKI M.

XX Erickson S, Schwall R, Sliwkowski M;

XX WPI; 2002-163686/21.

DR N-PSDB; ABK14058.

XX Treating tumour characterised by overexpression of epidermal growth
 FT factor receptor, ErbB or cancer in mammal, comprises administering
 FT anti-ErbB antibody-maytansinoid conjugate to the mammal -

XX Example 3; Fig 7; 93pp; English.

PS The invention relates to treating a tumour in a mammal, where the tumour

XX

CC is characterised by the overexpression of an epidermal growth factor
 CC receptor (ErbB) and does not respond or responds poorly, to treatment
 CC with an anti-ErbB antibody, comprising administering to the mammal an
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
 CC prostate and bladder, preferably breast cancer. The breast cancer is a
 CC metastatic breast cancer or an aggressive form of metastatic breast
 CC cancer which overexpresses ErbB2. The method is also useful for treating
 CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
 CC epithelial, stromal, blastocoealic, inflammatory, angiogenic and
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)
 CC polypeptide of the invention.

XX SQ Sequence 1255 AA;

Query Match 98.98; Score 6735; DB 23; Length 1255;

Best Local Similarity 99.08; Pred. No. 0;

Matches 1242; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLALLPPGAASQVCTGTDKMLRLPASPETHDMLRHLYGQCQVQGNL 60
 Db 1 MELAALCRWGLLALLPPGAASQVCTGTDKMLRLPASPETHDMLRHLYGQCQVQGNL 60
 Qy 61 ELYVLPNASLSFLQDIQEVQVQYIKANSKFIGITELQRLRIVRGTLQFEDNYALAVLNG 120
 Db 61 ELYVLPNASLSFLQDIQEVQVQYIAHNOVRQVPLQRLRIVRGTLQFEDNYALAVLNG 120
 Qy 121 DPLNNTPTVGTASPGGLRELQLRLSLTEILKGGVLIQBNPOLCYQDTTLWKDIFHKNOLA 180
 Db 121 DPLNNTPTVGTASPGGLRELQLRLSLTEILKGGVLIQBNPOLCYQDTTLWKDIFHKNOLA 180
 Qy 181 LTLIDTNRSRACHPCSPMKGSRGWESSEDCOSLTRTVCAGGCARCKGLPTDCCHEQC 240
 Db 181 LTLIDTNRSRACHPCSPMKGSRGWESSEDCOSLTRTVCAGGCARCKGLPTDCCHEQC 240
 Qy 241 AAGTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Db 241 AAGTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Qy 301 YNYLSTDVGSCTLVCPHNOEVTAEQGTORCEKSPCARVCYGLGMEHLREVRVAVTSAN 360
 Db 301 YNYLSTDVGSCTLVCPHNOEVTAEQGTORCEKSPCARVCYGLGMEHLREVRVAVTSAN 360
 Qy 361 IQBFAGCKKIFGSLAFLEPESFDGDPASNTAPLOEQLOVFETLEEITGYLYISAWPDSLP 420
 Db 361 IQBFAGCKKIFGSLAFLEPESFDGDPASNTAPLOEQLOVFETLEEITGYLYISAWPDSLP 420
 Qy 421 DLSVFQNLQVIRGRIHLNGAYSILTQGLGISWLSRLSRLSGLALIHNTLHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRIHLNGAYSILTQGLGISWLSRLSRLSGLALIHNTLHLCFVHTV 480
 Qy 481 PWDQLFNPHQALLHTANRPEDECVBEGGLACHQICARGHCWGPGTQVCNCSQPLRQBC 540
 Db 481 PWDQLFNPHQALLHTANRPEDECVBEGGLACHQICARGHCWGPGTQVCNCSQPLRQBC 540
 Qy 541 VEECRVLQGLFREYVVARHCLPCHPECPONGSVTCFPEADOCVACHYKDPFCVARC 600
 Db 541 VEECRVLQGLFREYVVARHCLPCHPECPONGSVTCFPEADOCVACHYKDPFCVARC 600
 Qy 601 PSGVKPDLSPYMPKPFDEEGACQPCPNCTHSCVDLDDKGCFAEQRASPLTSIVSAVG 660
 Db 601 PSGVKPDLSPYMPKPFDEEGACQPCPNCTHSCVDLDDKGCFAEQRASPLTSIVSAVG 660
 Qy 661 ILLVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
 Db 661 ILLVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
 Qy 721 RKVKVLGSGAFGTVYKGIWIPDGENVKI PVAIKVLRENTS PKANKETLDEAYVMAGVGP 780
 Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKI PVAIKVLRENTS PKANKETLDEAYVMAGVGP 780
 Qy 781 YVSRLLGICLTSTVQLVTQLMPIYGLLDHVRNRLGSGQDLNWCQIAKGMSYLEDVR 840

Db 781 YVSRLLGICLTSTVQLVTQMPYGCCLLDHVRNRRGLSGQDLNWCMIKAGMSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKMALESILRRFT 900
Qy 901 HQSDVWSYGVTVWELMTGAKPYDGIIPARETPDLEKGERLPQPPICITIDVYIMVKCWM 960
Db 901 HQSDVWSYGVTVWELMTGAKPYDGIIPARETPDLEKGERLPQPPICITIDVYIMVKCWM 960
Qy 961 IDSECRPRFRELVSFEFSRMARDPQRFVVIQNEDELGPASPLDSTFYRSILLEDDMGDLVDA 1020
Db 961 IDSECRPRFRELVSFEFSRMARDPQRFVVIQNEDELGPASPLDSTFYRSILLEDDMGDLVDA 1020
Qy 1021 EYLVPQGGFCPPDAPAGAGWVHRHRSSTRSGGDLTGLSPSEBEAPRPLAPSEG 1080
Db 1021 EYLVPQGGFCPPDAPAGAGWVHRHRSSTRSGGDLTGLSPSEBEAPRPLAPSEG 1080
Qy 1081 AGSDVFDGLGWAAGLQSLPTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGLGWAAGLQSLPTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NQPDVROPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPO 1200
Db 1141 NQPDVROPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAFSPAFNLVYWDQPPERGAPSTFKGTPTAENPEYGLDVPV 1255
Db 1201 GGAAPQHPHPPAFSPAFNLVYWDQPPERGAPSTFKGTPTAENPEYGLDVPV 1255

RESULT 5

AAW01111 standard; Protein; 1255 AA.

AC AAW01111;

DT 01-JAN-1997 (first entry)

DE HER-2/neu protein.

KW HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW breast cancer; ovary cancer; colon cancer; lung cancer;
KW prostate cancer; immunisation; tumour; vaccine; vector.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Domain 676..1255

FT /label= Intracellular_domain

FT /note= "claimed domain, useful for immunisation"

XX W09630514-A1.

XX 03-OCT-1996.

XX 28-MAR-1996; 96WO-US01689.

XX 31-MAR-1995; 95US-0414417.

XX (UNIW) UNIV WASHINGTON.

XX Cheever MA, Disis ML;

XX WPI; 1996-455361/45.

XX N-PSDB; AAT40739.

PT DNA encoding HER-2-neu poly-peptide(s) - used for prevention or
PT treatment of malignancies with which the HER-2/neu oncogene is
XX associated

PS Claim 2; Page 56-61; 71pp; English.

XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal calls are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.

XX SQ Sequence 1255 AA;

Query Match 98.8%; Score 6729; DB 17; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;

Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPETHLDMLRHLVGGCQVQGNL 60

Db 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPETHLDMLRHLVGGCQVQGNL 60

Qy 61 ELYLPTNASLSFLQDIQEVQYVLIHNNQVRQVPLQLRLIVRGTLFEDNYALAVLNG 120

Db 61 ELYLPTNASLSFLQDIQEVQYVLIHNNQVRQVPLQLRLIVRGTLFEDNYALAVLNG 120

Qy 121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPOLCYODTILWKDI PHKNQOLA 180

Db 121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPOLCYODTILWKDI PHKNQOLA 180

Qy 181 LTLIDTNRSRACHPCSPMKGRCWGSESDCQSLTRTVCAAGCARCKGPLPTDCCHEQC 240

Db 181 LTLIDTNRSRACHPCSPMKGRCWGSESDCQSLTRTVCAAGCARCKGPLPTDCCHEQC 240

Qy 241 AAGCTGPKHSDCLACHFNHSGICELCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Db 241 AAGCTGPKHSDCLACHFNHSGICELCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy 301 YNYLSTDVGSCTLVCPLNHNOEVTAEADGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360

Db 301 YNYLSTDVGSCTLVCPLNHNOEVTAEADGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360

Qy 361 IQEFAGCKKIFGSLAFPEFSDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSL 420

Db 361 IQEFAGCKKIFGSLAFPEFSDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSL 420

Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWLGRLSRLGSLALIHNNTHLCFVHTV 480

Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWLGRLSRLGSLALIHNNTHLCFVHTV 480

Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy 541 VEECRVLQGLPREYVYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600

Db 541 VEECRVLQGLPREYVYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600

Qy 601 PSGVKPDLSPYIWKFPDEEGACQPCPINCTHSCVDLDDKGCAPAEORASPLTSIVSAVVG 660

Db 601 PSGVKPDLSPYIWKFPDEEGACQPCPINCTHSCVDLDDKGCAPAEORASPLTSIVSAVVG 660

Qy 661 ILLVVLGVVFGILIKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Db 661 ILLVVLGVVFGILIKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 780

Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 780

Qy 781 YVSRLLGICLTSTVQLVTQMPYGCCLLDHVRNRRGLSGQDLNWCMIKAGMSYLEDVR 840

Db 781 YVSRLLGICLTSTVQLVTQMPYGCCLLDHVRNRRGLSGQDLNWCMIKAGMSYLEDVR 840

QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900
 DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900
 QY 901 HQSDVWSYGVTVWELMTFGAKPYDGI PARETPDLLEKGERLPQPPICITIDVYIMVKCWM 960
 DB 901 HQSDVWSYGVTVWELMTFGAKPYDGI PARETPDLLEKGERLPQPPICITIDVYIMVKCWM 960
 QY 961 IDSECRPRFRELSEFSEFARMARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDGMDLVD 1020
 DB 961 IDSECRPRFRELSEFSEFARMARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDGMDLVD 1020
 QY 1021 EYILVPOQGFPCPDPAFCAGMWHRRSSSTRSGGDLTLGLPSEBAPRSPASEG 1080
 DB 1021 EYILVPOQGFPCPDPAFCAGMWHRRSSSTRSGGDLTLGLPSEBAPRSPASEG 1080
 QY 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLORYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
 DB 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLORYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
 QY 1141 NQPDVREPQPSREGPLPAARPGATLERAKTSLSPGKNGVVKDVFAGGAVENPEYLTPQ 1200
 DB 1141 NQPDVREPQPSREGPLPAARPGATLERAKTSLSPGKNGVVKDVFAGGAVENPEYLTPQ 1200
 QY 1201 GGAAPQHPHPPAFSPADNLYYWDQDPPERGAPSTFKGTPTAENPEYGLDVPV 1255
 DB 1201 GGAAPQHPHPPAFSPADNLYYWDQDPPERGAPSTFKGTPTAENPEYGLDVPV 1255

RESULT 6

AAW92406
 ID AAW92406 standard; Protein; 1255 AA.

AAW92406;

21-APR-1999 (first entry)

Human HER-2/neu oncogene protein.

HER-2/neu; oncogene; immune response; T cell; B cell; immunisation; malignancy; treatment; tumour.

Homo sapiens.

Key Location/Qualifiers

Region 676..1255

/note= "region which elicits immune response"

US5869445-A.

09-FEB-1999.

01-APR-1996; 96US-0625101.

01-APR-1996; 96US-0625101.

17-MAR-1993; 93US-0033644.

12-AUG-1993; 93US-0106112.

31-MAR-1995; 95US-0414417.

(UNIW) UNIV WASHINGTON.

Cheever MA, Disis ML;

WPI; 1999-152835/13.

N-PSDB; AAX01912.

Use of HER-2/neu polypeptides - for eliciting an immune response to an HER-2/neu associated malignancy, particularly for treating or preventing tumours

Claim 3; Column 31-38; 26pp; English.

This sequence represents the human HER-2/neu oncogene protein. A fragment

CC of this protein is used in a method for eliciting or enhancing an immune response to HER-2/neu protein. The polypeptide can stimulate T cells and B cells to produce an immune response to the HER-2/neu protein. The method can be used for immunisation against a malignancy in which the HER-2/neu oncogene is associated and in the treatment of an existing tumour, or to prevent tumour occurrence or reoccurrence.

XX Sequence 1255 AA;

Query Match 98.8%; Score 6729; DB 20; Length 1255;
 Best Local Similarity 98.8%; Pred. No. 0;
 Matches 1240; Conservative. 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALLPPGAASQVCTGTDMLKRLPASPTHLDMLRHLYGGCQVQGNL 60
 DB 1 MELAALCRWGLLALLPPGAASQVCTGTDMLKRLPASPTHLDMLRHLYGGCQVQGNL 60
 QY 61 ELYLPTNASLSFLDIOEQVQYIKANSKFIGITELORLIRVGTQLFEDNYALAVLDNG 120
 DB 61 ELYLPTNASLSFLDIOEQVQYIKANSKFIGITELORLIRVGTQLFEDNYALAVLDNG 120
 QY 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYODTILWKDIFHKNNOLA 180
 DB 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYODTILWKDIFHKNNOLA 180
 QY 181 LTLIDTNRSRACHPCSPMCKGSRGSESSDCOSLTRTVCCAGCARCKGPLPTDCCHEQC 240
 DB 181 LTLIDTNRSRACHPCSPMCKGSRGSESSDCOSLTRTVCCAGCARCKGPLPTDCCHEQC 240
 QY 241 AAGCTGPKGISDCLACLHFNHSGICELHCPALVYNTDTPFESMPNPEGRTYFGASCVTACP 300
 DB 241 AAGCTGPKGISDCLACLHFNHSGICELHCPALVYNTDTPFESMPNPEGRTYFGASCVTACP 300
 QY 301 YNYLSTDVGSCTLVCPPLHQEVTAEDGTORCEKSCPCARVCYGLGMEHLREVRVTSAN 360
 DB 301 YNYLSTDVGSCTLVCPPLHQEVTAEDGTORCEKSCPCARVCYGLGMEHLREVRVTSAN 360
 QY 361 IQEFAGCKIIFGSLAFPLPESFDGDPASNTAPLOEQLOVFETLEEITGYLIYISAWPDSL 420
 DB 361 IQEFAGCKIIFGSLAFPLPESFDGDPASNTAPLOEQLOVFETLEEITGYLIYISAWPDSL 420
 QY 421 DLSVFQNLQVIRGRIILHNGAYSILTLQGLISWIGLSRLGSLALIHNTLHLCFVHTV 480
 DB 421 DLSVFQNLQVIRGRIILHNGAYSILTLQGLISWIGLSRLGSLALIHNTLHLCFVHTV 480
 QY 481 PDQLFRNPHQALLHTANRPEDECVGEGACHOLCARGHCWGPGTQCVCNCSQFLRGQEC 540
 DB 481 PDQLFRNPHQALLHTANRPEDECVGEGACHOLCARGHCWGPGTQCVCNCSQFLRGQEC 540
 QY 541 VEECRVLQGLFREYVYVNAHCLPCHPECQPNQSVTCFGEADQCVACAHYKDPFVCVARC 600
 DB 541 VEECRVLQGLFREYVYVNAHCLPCHPECQPNQSVTCFGEADQCVACAHYKDPFVCVARC 600
 QY 601 PSGVKPDLSPYIWKFPDSEBACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVG 660
 DB 601 PSGVKPDLSPYIWKFPDSEBACQPCPINCTHSCVDLDDKGCPAEQRASPLTSIVSAVG 660
 QY 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOTELVPELTPSGAMPNQAMRILKETEL 720
 DB 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOTELVPELTPSGAMPNQAMRILKETEL 720
 QY 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780
 DB 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780
 QY 781 YVSRLLGICLTSTVQLVLTQMPYVGLLDHVRENRLGSGDQLLNWCWQIAKMSYLEDVR 840
 DB 781 YVSRLLGICLTSTVQLVLTQMPYVGLLDHVRENRLGSGDQLLNWCWQIAKMSYLEDVR 840
 QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900
 DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900

Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQPPICITIDVYIMVWKCWM 960
 Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQPPICITIDVYIMVWKCWM 960
 Qy 961 IDSECRPRFRFELVSEFSRMDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDDMGDLVDA 1020
 Db 961 IDSECRPRFRFELVSEFSRMDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDDMGDLVDA 1020
 Qy 1021 BEYLVPQGFPCPDPAFCAGGCVHRRHSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080
 Db 1021 BEYLVPQGFPCPDPAFCAGGCVHRRHSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080
 Qy 1081 AGSDVFDGLGWAAGKLOSLPTDPSPLORYSDPTVPLPSETDGVVAPLTCSPQPEYV 1140
 Db 1081 AGSDVFDGLGWAAGKLOSLPTDPSPLORYSDPTVPLPSETDGVVAPLTCSPQPEYV 1140
 Qy 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTLPFGKNGVVKDVFAGGAVENPEYLTPQ 1200
 Db 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTLPFGKNGVVKDVFAGGAVENPEYLTPQ 1200
 Qy 1201 GGAAPQHPHPPAFSPADNLVYWDQPPPERGAPSTFKGTTAENPEYGLDVPV 1255
 Db 1201 GGAAPQHPHPPAFSPADNLVYWDQPPPERGAPSTFKGTTAENPEYGLDVPV 1255

RESULT 7

AAB21198
 ID AAB21198 standard; protein; 1255 AA.

AC AAB21198;

DT 12-JAN-2001 (first entry)

XX Human HER-2/neu protein.

XX Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
 KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
 KW colon cancer.

OS Homo sapiens.

PN WO200044899-A1.

PD 03-AUG-2000.

PF 28-JAN-2000; 2000WO-US02164.

PR 29-JAN-1999; 99US-0117976.

XX (CORI-) CORIXA CORP.
 PA (SMIK-) SMITHKLINE BEECHAM.

XX Cheever MA, Gheysen D;

XX WPI; 2000-505976/45.

XX N-PSDB; AAA89736.

PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins
 PT useful for vaccinating against breast, ovarian, colon, lung and
 PT prostate cancers -

XX Claim 52; Fig 7; 128pp; English.

XX The present sequence is the human HER-2/neu protein. It is a member of
 CC the tyrosine kinase family of receptor-like glycoproteins and shows
 CC homology to the epidermal growth factor receptor (EGFR). It probably
 CC plays a part in cell growth and/or differentiation. The HER-2/neu
 CC gene is an oncogene. An HER-2/neu fusion protein comprising a
 CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
 CC domain may be used to treat or prevent cancer by eliciting or
 CC enhancing an immune response to the HER-2/neu protein. It may be used
 CC to treat malignancies such as breast, ovarian, colon, lung and
 CC prostate, cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.
 XX
 SQ Sequence 1255 AA;

Query Match 98.8%; Score 6729; DB 21; Length 1255;
 Best Local Similarity 98.8%; Pred. No. 0;
 Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLALLPPGAASSTOCTGTDMKRLRLPASPTHLDMLRHLHQGVVQGNL 60
 Db 1 MELAALCRWGLLALLPPGAASSTOCTGTDMKRLRLPASPTHLDMLRHLHQGVVQGNL 60
 Qy 61 ELYVLPNASLSFLQDIQEVQVQYIKANSKEFIGITELQRLAIRVGTQLFEDNYALAVLDNG 120
 Db 61 ELYVLPNASLSFLQDIQEVQVQYIKANSKEFIGITELQRLAIRVGTQLFEDNYALAVLDNG 120
 Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNQLA 180
 Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNQLA 180
 Qy 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLTRTVCCAGCARCKGPLPTDCCHEQC 240
 Db 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLTRTVCCAGCARCKGPLPTDCCHEQC 240
 Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Qy 301 YNYLSTDVGSCTLVCPLNHNOEVTAEQGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
 Db 301 YNYLSTDVGSCTLVCPLNHNOEVTAEQGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
 Qy 361 IQEFAGCKKIFGSLAFPEFSDGDPASNTAPLOEQLOVFETLEEITGYLYISAWPSLP 420
 Db 361 IQEFAGCKKIFGSLAFPEFSDGDPASNTAPLOEQLOVFETLEEITGYLYISAWPSLP 420
 Qy 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLGSLALIHNTLHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLGSLALIHNTLHLCFVHTV 480
 Qy 481 PWDQLFNPHQALLHTANRPEDCEVGEGLACHQICARGHCWGPGPTQCVNCSQPLRGQEC 540
 Db 481 PWDQLFNPHQALLHTANRPEDCEVGEGLACHQICARGHCWGPGPTQCVNCSQPLRGQEC 540
 Qy 541 VESCRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFCVARC 600
 Db 541 VESCRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFCVARC 600
 Qy 601 PSGVKPDLSYNPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSISAVVG 660
 Db 601 PSGVKPDLSYNPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSISAVVG 660
 Qy 661 ILLVVLGVVFGILIKRRQKIRKYTNRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
 Db 661 ILLVVLGVVFGILIKRRQKIRKYTNRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
 Qy 721 RKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780
 Db 721 RKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780
 Qy 781 YVSRLLGICLTSTVOLVTQLMYPGCLLDHVRNRLGRLGSDLLNWCMIKAGMSYLEDVR 840
 Db 781 YVSRLLGICLTSTVOLVTQLMYPGCLLDHVRNRLGRLGSDLLNWCMIKAGMSYLEDVR 840
 Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILRRFT 900
 Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILRRFT 900
 Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQPPICITIDVYIMVWKCWM 960
 Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQPPICITIDVYIMVWKCWM 960
 Qy 961 IDSECRPRFRFELVSEFSRMDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDDMGDLVDA 1020

Db 961 IDSECRFRRLVSEFRSMARDPQRFVVIQNEIDGPASPLDSTFYRSLLEDDDDGLVDA 1020
Qy 1021 BEYLVPQGGFFCPDPAFCAGGVMHRRSSSTRSGGDLTLGLPSPSEEAAPRSLAPSEG 1080
Db 1021 BEYLVPQGGFFCPDPAFCAGGVMHRRSSSTRSGGDLTLGLPSPSEEAAPRSLAPSEG 1080
Qy 1081 AGSDVFGDGLGMAAGLQSLPHTDPSPLQRYSDPTVPLPSETDGVYAPLTCSPQPEYV 1140
Db 1081 AGSDVFGDGLGMAAGLQSLPHTDPSPLQRYSDPTVPLPSETDGVYAPLTCSPQPEYV 1140
Qy 1141 NQDVRPQPPSPRGPLPAARPAGATLERAKTLPSPKNGVVKDYFAFGGAVENPEYLTPO 1200
Db 1141 NQDVRPQPPSPRGPLPAARPAGATLERAKTLPSPKNGVVKDYFAFGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAFSPADNLYYWDPPPERGAPPTFKGTPTAENPEYLGIDVPV 1255
Db 1201 GGAAPQHPHPPAFSPADNLYYWDPPPERGAPPTFKGTPTAENPEYLGIDVPV 1255

RESULT 8
AAY84780
ID AAY84780 standard; Protein; 1255 AA.
XX
AC AAY84780;
XX
AC
XX
DT 08-AUG-2000 (first entry)
XX
DE Amino acid sequence of the SPLICE erbB-2 receptor protein.
XX
KW SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;
KW tumor cell proliferation; tissue degeneration; arthropathy;
KW bone resorption; inflammatory disease; degenerative disorder;
KW wound healing.
XX
XX Homo sapiens.
XX
XX W0200020579-Al.
XX
XX 13-APR-2000.
XX
XX 01-OCT-1999; 99WO-CA00912.
XX
XX 02-OCT-1998; 98US-0165192.
XX
XX (UYMC-) UNIV MCMASTER.
XX
XX Muller WJ, Siegel PM;
XX
XX WPI; 2000-303768/26.
XX
XX N-PSDB; AAA14812.
XX
XX Nucleic acid encoding an erbB 2 receptor protein designated SPLICE
XX erbB-2, inhibitors of the protein are useful for treatment of cancer -
XX
XX Claim 3; Fig 2; 60pp; English.
XX
XX The present sequence represents a SPLICE erbB-2 receptor protein. The
XX protein has an in-frame deletion of 16 amino acids, 2 of which are
XX conserved cysteine residues, compared to the unspliced protein. The
XX erbB-2 polynucleotide is used to construct probes for detecting
XX disorders of cell transformation such as cancer. Antibodies to the
XX protein may be used to detect SPLICE erbB-2 in a sample. Agents
XX (e.g. antisense oligonucleotides) which inhibit the expression of
XX SPLICE erbB-2 are useful for reducing tumor cell proliferation and
XX treating cancer. Substances which stimulate SPLICE erbB-2 are useful
XX for treating conditions involving damaged cells including conditions
XX in which degeneration of tissue occurs, such as arthropathy, bone
XX resorption, inflammatory diseases, degenerative disorders of the
XX central nervous system and wound healing.
XX
XX Sequence 1255 AA;

Query Match 98.8%; Score 6729; DB 21; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;
Qy 1 MEAALCRWGLLALLPPGAASQVCTGTDMKRLPASPTHLDMLRHLVGGCVOVGNL 60
Db 1 MEAALCRWGLLALLPPGAASQVCTGTDMKRLPASPTHLDMLRHLVGGCVOVGNL 60
Qy 61 ELTYLPTNASLSLQDIOEQVQYIKANSKFIQITELQRLRIVRGTLQFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSLQDIOEQVQYIKANSKFIQITELQRLRIVRGTLQFEDNYALAVLDNG 120
Qy 121 DPLNNTPTVGTASPGGLRELQLRSLTEILKGVLIQRNPOLCYODTTLWKDIFHKNOLA 180
Db 121 DPLNNTPTVGTASPGGLRELQLRSLTEILKGVLIQRNPOLCYODTTLWKDIFHKNOLA 180
Qy 181 LTLIDNRSRACHPCSPMKSGSCWGESSEDCSLTRTVCCAGGCARCKGLPTDCCHEQC 240
Db 181 LTLIDNRSRACHPCSPMKSGSCWGESSEDCSLTRTVCCAGGCARCKGLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPLHNOEVTABDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPLHNOEVTABDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IQBFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEETIGYLIYISAWPSLP 420
Db 361 IQBFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEETIGYLIYISAWPSLP 420
Qy 421 DLSVFQNLQVIRGRIHLHNGAYSILTLOGLGISWLSGLRSLRELGLALIHNTLHCFVHTV 480
Db 421 DLSVFQNLQVIRGRIHLHNGAYSILTLOGLGISWLSGLRSLRELGLALIHNTLHCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPGTQCVNCSQFLRQOEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPGTQCVNCSQFLRQOEC 540
Qy 541 VEESRVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACHYKDPFCVARC 600
Db 541 VEESRVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACHYKDPFCVARC 600
Qy 601 PSGVKPDLSPYMPIWKPPDEGACQPCINCHSCVDLDDKGCPCAEQASPLTSTVSAVVG 660
Db 601 PSGVKPDLSPYMPIWKPPDEGACQPCINCHSCVDLDDKGCPCAEQASPLTSTVSAVVG 660
Qy 661 ILLVWVGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVWVGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVSP 780
Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVSP 780
Qy 781 YVSRLLGICLTSTVQLVTQLMYPYGLLDHVRNRLGSDLLNWCMIKAGMSYLDVR 840
Db 781 YVSRLLGICLTSTVQLVTQLMYPYGLLDHVRNRLGSDLLNWCMIKAGMSYLDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLDDIDETEHADGGKVPIKWMALLESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLDDIDETEHADGGKVPIKWMALLESILRRFT 900
Qy 901 HQSDVWSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPCTIDVYIMWKCMW 960
Db 901 HQSDVWSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPCTIDVYIMWKCMW 960
Qy 961 IDSECRPRFRELVSFESRWARDPQRFVVIQNEIDGPASPLDSTFYRSLLEDDDDGLVDA 1020
Db 961 IDSECRPRFRELVSFESRWARDPQRFVVIQNEIDGPASPLDSTFYRSLLEDDDDGLVDA 1020
Qy 1021 EBYLVFPQGGFFCPDPAFCAGGVMHRRSSSTRSGGDLTLGLPSPSEEAAPRSLAPSEG 1080

```
Db 1021 EEYLVPOQGFPCDPAPGAGGVHRRSSSTRSGGDLTLGLPSEEA PRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPQPSPRGPIPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTPO 1200
Db 1141 NOPDVRPQPSPRGPIPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPAPFADNLYWDPPERGAPSTFKGTPTAENPEYLGIDVFPV 1255
Db 1201 GGAAPQHPHPAPFADNLYWDPPERGAPSTFKGTPTAENPEYLGIDVFPV 1255

RESULT 9
AAB85458
ID AAB85458 standard; Protein; 1255 AA.
XX
AC AAB85458;
XX
DT 25-SEP-2001 (first entry)
XX
DE Human HER-2/neu protein.
XX
KW Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KW oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX
OS Homo sapiens.
XX
XX WO200153463-A2.
XX
PD 26-JUL-2001.
XX
XX 19-JAN-2001; 2001WO-US01850.
XX
XX 21-JAN-2000; 2000US-0177545.
XX
XX (CORI-) CORIXA CORP.
XX
XX Cheever MA, Hand-Zimmermann S;
XX
XX WPI; 2001-476112/51.
XX
XX N-PSDB; AAH23392.
XX
XX New antigen-presenting cells, useful as vaccines for eliciting or
XX enhancing an immune response to HER-2/neu protein, particularly useful
XX for treating or preventing cancer, e.g. breast cancer -
XX
XX Claim 2; Page 41-46; 49pp; English.
XX
XX The invention provides an isolated antigen-presenting cell, which
XX expresses at least an immunogenic portion of a polypeptide that produces
XX an immune response to HER-2/neu protein. The antigen-presenting cells are
XX useful as vaccines for eliciting or enhancing an immune response to
XX HER-2/neu protein, particularly in treating or preventing malignancies in
XX which the HER-2/neu oncogene is associated. Specifically, these are
XX useful for treating or preventing cancer, e.g. breast cancer, ovarian,
XX colon, lung or prostate cancers. The present sequence represents
XX the human HER-2/neu protein (also known as p185 or c-erbB2).
XX
XX Sequence 1255 AA;
XX
XX Query Match 98.8%; Score 6729; DB 22; Length 1255;
XX Best Local Similarity 98.8%; Pred. NO. 0;
XX Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;
XX
XX 1 MELAAALCRWGLLALLPPGAASQTCTGDMKRLPASPETHLDMRLHYGCGVQGNL 60
XX 1 MELAAALCRWGLLALLPPGAASQTCTGDMKRLPASPETHLDMRLHYGCGVQGNL 60
XX
XX 61.ELTYLPTNASLSFLQDIQEVQVYIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
```

```
Db 61 ELTYLPTNASLSFLQDIQEVQVYIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIHKNOLA 180
Db 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIHKNOLA 180
Qy 181 LTLIDNRSRACHPCSPMKCKSGSCWGESSEDCOSLRTVTCAGGCARCKGLPTDCCHEQC 240
Db 181 LTLIDNRSRACHPCSPMKCKSGSCWGESSEDCOSLRTVTCAGGCARCKGLPTDCCHEQC 240
Qy 241 AAGCTGPKGSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKGSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHREVRVAVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHREVRVAVTSAN 360
Qy 361 IQEFAGCKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEEITGYLYISAWPDSL 420
Db 361 IQEFAGCKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEEITGYLYISAWPDSL 420
Qy 421 DLSVFQNLQVIRGRIHLHGAYSLTLQGLGISWLGSLRSLRELGSGLALIHNTLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRIHLHGAYSLTLQGLGISWLGSLRSLRELGSGLALIHNTLHLCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRDECEVGEGLACHQLCARGHCWGPGTQVCNCSQFLRQOEC 540
Db 481 PWDQLFRNPHQALLHTANRDECEVGEGLACHQLCARGHCWGPGTQVCNCSQFLRQOEC 540
Qy 541 VECRVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFVCVARC 600
Db 541 VECRVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFVCVARC 600
Qy 601 PSGVKPDLSTYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCPEAQSPASLTISAVVG 660
Db 601 PSGVKPDLSTYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCPEAQSPASLTISAVVG 660
Qy 661 ILLVVLGVVFGILIKRQOKIRKYTWRLLOETELVEPLTPSGAMNQAMRLKETE 720
Db 661 ILLVVLGVVFGILIKRQOKIRKYTWRLLOETELVEPLTPSGAMNQAMRLKETE 720
Qy 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKELDEAYVWAGVSP 780
Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKELDEAYVWAGVSP 780
Qy 781 YVSRLLGICLTSTVQLVTQLMFYCCLLDHVRENRLGSLQDLNWCQIAKMSYLEEDVR 840
Db 781 YVSRLLGICLTSTVQLVTQLMFYCCLLDHVRENRLGSLQDLNWCQIAKMSYLEEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLDDIDETEHADGGKVPDKWMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLDDIDETEHADGGKVPDKWMALESILRRFT 900
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIPIREIPDLLEKGERLPOPICTIDVYMLWKCWM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIPIREIPDLLEKGERLPOPICTIDVYMLWKCWM 960
Qy 961 IDSECRPRFRELVESEFSRMDARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDDGLVDA 1020
Db 961 IDSECRPRFRELVESEFSRMDARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDDGLVDA 1020
Qy 1021 EYILVPOQGFPCDPAPGAGGVHRRSSSTRSGGDLTLGLPSEEA PRSLAPSEG 1080
Db 1021 EYILVPOQGFPCDPAPGAGGVHRRSSSTRSGGDLTLGLPSEEA PRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPQPSPRGPIPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTPO 1200
```

Db 1141 NQDVRPQPPSPREGPLPAARPAGATLERPKTSLSPGKNGVVDYFAFGGAVENPEYLTPO 1200
Qy 1201 GGAAPOHPHPPAFSPAFDNLYWDQDPPERGAPSTFKGTPTAENPEYLGLDVDFV 1255
Db 1201 GGAAPOHPHPPAFSPAFDNLYWDQDPPERGAPSTFKGTPTAENPEYLGLDVDFV 1255
RESULT 10
AAG88267
ID AAG88267 standard; Protein; 1255 AA.
XX
AC AAG88267;
XX
DT 11-SEP-2001 (first entry)
XX
DE HER2/neu amino acid sequence.
XX
KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
OS Homo sapiens.
XX
PN WO200141787-A1.
XX
PD 14-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US33591.
XX
PR 10-DEC-1999; 99US-0458299.
XX
PA (EPIM-) EPIMUNE INC.
XX
PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
PI Keogh E;
XX
DR WPI; 2001-374995/39.
XX
PT An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer -
PS Disclosure; Page 15; 199pp; English.
XX
CC The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
CC and immunostimulant activities, and can be used in vaccines. (I), (II)
CC and (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample from a patient and
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention.
XX
SQ Sequence 1255 AA;

Query Match

98.8%; Score 6729; DB 22; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLALLPFGAASQVCTGTDKMLRLPASPTHLDMLRHLHGCGOVQGNL 60
Db 1 MELAALCRWGLLALLPFGAASQVCTGTDKMLRLPASPTHLDMLRHLHGCGOVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQVYVLIHAHQVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQVYVLIHAHQVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTTLWKDIHFNKNOLA 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTTLWKDIHFNKNOLA 180
Qy 181 LTLIDTNRSRACHPCSPMKCKGSRGWESSDCOSLRTVTCAGGCARCKGLPTDCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMKCKGSRGWESSDCOSLRTVTCAGGCARCKGLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPPLHNOEVTAEADGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPPLHNOEVTAEADGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IQEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOEQVETLEEITGYLYISAMPDSLP 420
Db 361 IQEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOEQVETLEEITGYLYISAMPDSLP 420
Qy 421 DLSVFQNLQVIRGRIHLNGAYSILTQGLGISWLSRLSRLSGLSGLALIHNTLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRIHLNGAYSILTQGLGISWLSRLSRLSGLSGLALIHNTLHLCFVHTV 480
Qy 481 PWDLFRNPHQALLHTANRPEDECEVGEGLACHOLCARGHCWGPGTQVCNCSOFLRQEC 540
Db 481 PWDLFRNPHQALLHTANRPEDECEVGEGLACHOLCARGHCWGPGTQVCNCSOFLRQEC 540
Qy 541 VESCRVLQGLPREYVNAHCLPCHPFCOPONGSVTCFGEADOCVCAHVKDPPFCVARC 600
Db 541 VESCRVLQGLPREYVNAHCLPCHPFCOPONGSVTCFGEADOCVCAHVKDPPFCVARC 600
Qy 601 PSGVKPDLSTYMPKPFDEEGACQPCPINCTHSCVDLDDKGCAPAEORASPLTSIVSAVG 660
Db 601 PSGVKPDLSTYMPKPFDEEGACQPCPINCTHSCVDLDDKGCAPAEORASPLTSIVSAVG 660
Qy 661 ILLVWVLGVVFGILIKRRQOKIRKYTMRRLLQSTELVEPLTPSGAMPNQAOMRILKETEL 720
Db 661 ILLVWVLGVVFGILIKRRQOKIRKYTMRRLLQSTELVEPLTPSGAMPNQAOMRILKETEL 720
Qy 721 RKVKVLGSGAFGVYKGIWTPDGENVKIPVAIKVLRENTSPKANKETLDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGVYKGIWTPDGENVKIPVAIKVLRENTSPKANKETLDEAYVMAGVGP 780
Qy 781 YVSRLLGICLTSTVQLVTLQMPYGCLLDHVRENRLGSDQLLNCWQIAKGSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTLQMPYGCLLDHVRENRLGSDQLLNCWQIAKGSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLIDDETYHADGGKVPKWMALLESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLIDDETYHADGGKVPKWMALLESILRRFT 900
Qy 901 HQSDVMSGYVWVWELMTFGAKPYDGIIPAREIPOLLEKGERLPPOPCTIDVYIMVKCWM 960
Db 901 HQSDVMSGYVWVWELMTFGAKPYDGIIPAREIPOLLEKGERLPPOPCTIDVYIMVKCWM 960
Qy 961 IDSECRPRFRELVSERFARMARDPQRFVWQNEIDGLPASPLDSTFYRSLLEDGMDLVDA 1020
Db 961 IDSECRPRFRELVSERFARMARDPQRFVWQNEIDGLPASPLDSTFYRSLLEDGMDLVDA 1020
Qy 1021 EYVLVQQGFPCDPPAPGAGGVMHRRSSSTSGGGDLTLGLERPEEAPRSLAPSEG 1080
Db 1021 EYVLVQQGFPCDPPAPGAGGVMHRRSSSTSGGGDLTLGLERPEEAPRSLAPSEG 1080

Db 1021 EYLVPQGFCDPAPGAGGVHRRSSSTRSGGDLTLGLPSEEA PRSPLAPSEG 1080
Qy 1081 AGSDVFGDLGMAAKGLQSLPTDHPSPLOYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFGDLGMAAKGLQSLPTDHPSPLOYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPQPPSPRGPLPAARPAAGATLERAKTLPSPKNGVVKDVPFAGGAVENPEYLTPO 1200
Db 1141 NOPDVRPQPPSPRGPLPAARPAAGATLERAKTLPSPKNGVVKDVPFAGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAFPAFNLYYWQDPPRGAPSTFKGTPTAENPEYLGLDVFP 1255
Db 1201 GGAAPQHPHPPAFPAFNLYYWQDPPRGAPSTFKGTPTAENPEYLGLDVFP 1255

RESULT 11

AAE24067
ID AAE24067 standard; Protein; 1255 AA.
XX
AC AAE24067;
XX
XX 23-SEP-2002 (first entry)
XX
DE Human Her-2 protein.
XX
KW Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KW hyperproliferative disorder; prophylaxis; inflammation; antisense;
KW tumour; gene therapy; phosphorothioate backbone.
XX
OS Homo sapiens.
XX
PN WO200022636-A1.
XX
PD 21-MAR-2002.
XX
XX 12-SEP-2001; 2001WO-US28572.
XX
PR 15-SEP-2000; 2000US-0663834.
XX
XX (ISIS-) ISIS PHARM INC.
XX
PI Bennett CF, Cowse LM;
XX
XX WPI; 2002-471192/50.
DR N-PSDB; AAD38904.
XX
PT Novel antisense oligonucleotide which modulates the expression of Human
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT inflammation or to prevent infection in humans -
XX
PS Example 13; Page 95-107; 116pp; English.
XX

The invention relates to antisense compounds targeted to a nucleic acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2) that specifically hybridises with and inhibits the expression of Her2. Antisense compounds of the invention are used for treating diseases or conditions associated with Her2 such as hyperproliferative disorders e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary, neural or cardiac cancer. They are also useful prophylactically e.g. to prevent or delay infection, inflammation and tumour formation. The invention is also used in gene therapy. The present sequence is human Her-2 protein.

XX Sequence 1255 AA;

Query Match 98.8%; Score 6729; DB 23; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MELAACRWGLLLALLPPGAASTQVCTGDMKRLPASPETHDMLRLHYGQCQVQGNL 60
Db 1 MELAACRWGLLLALLPPGAASTQVCTGDMKRLPASPETHDMLRLHYGQCQVQGNL 60

Qy 61 ELTYLPTNASLSFLQDIOEQVQYIKANSKFIGITELQRLRIVRGTQOLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEQVQYIKANSKFIGITELQRLRIVRGTQOLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTTLWKDIFHKNQOLA 180
Db 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTTLWKDIFHKNQOLA 180
Qy 181 LTLIDNRSRACHPCSPMKGRCWGSSEDCOSLRTFVCAGGCARCKGPLPTDCCHEQC 240
Db 181 LTLIDNRSRACHPCSPMKGRCWGSSEDCOSLRTFVCAGGCARCKGPLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPPLHNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPPLHNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IOBFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEEITGYLYISAMPDLSL 420
Db 361 IOBFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEEITGYLYISAMPDLSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLGSLALIHNTLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLGSLALIHNTLHLCFVHTV 480
Qy 481 PWDQFRNPHQALLHTANRPEDECVGEGLAHQICARGHCHWGPGPTQCVNCSQFLRQEC 540
Db 481 PWDQFRNPHQALLHTANRPEDECVGEGLAHQICARGHCHWGPGPTQCVNCSQFLRQEC 540
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFCVARC 600
Qy 601 PSGVKPDLSPYMPWKFPDEEGACQPCINCTHSCVDLDDKGCAPAEQASPLTSIVSAVVG 660
Db 601 PSGVKPDLSPYMPWKFPDEEGACQPCINCTHSCVDLDDKGCAPAEQASPLTSIVSAVVG 660
Qy 661 ILLVVLGVVFGILIKRRQKIRKYTWRLQLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVLGVVFGILIKRRQKIRKYTWRLQLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Qy 721 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRNTSPKANKETILDEAYVWAGVSP 780
Db 721 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRNTSPKANKETILDEAYVWAGVSP 780
Qy 781 YVSRLLGICLTSTVOLVTLMPYGLDHDVRENRGLSQDLDLNCMOIAKMSYLEDVR 840
Db 781 YVSRLLGICLTSTVOLVTLMPYGLDHDVRENRGLSQDLDLNCMOIAKMSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDPLGLARLLIDETEHADGGKVPPIKMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDPLGLARLLIDETEHADGGKVPPIKMALESILRRFT 900
Qy 901 HQSDVWSYGVTVWELMTFGAPYDGIIPAREIPDLLEKGERLPQPPICITDVIYMIWKCWM 960
Db 901 HQSDVWSYGVTVWELMTFGAPYDGIIPAREIPDLLEKGERLPQPPICITDVIYMIWKCWM 960
Qy 961 IDSECRPRFRELVSFERSMARDPQFVVIQNEIDLGPASPLDSTFYRSILLEDDMGDLVDA 1020
Db 961 IDSECRPRFRELVSFERSMARDPQFVVIQNEIDLGPASPLDSTFYRSILLEDDMGDLVDA 1020
Qy 1021 EYLVPQGFCDPAPGAGGVHRRSSSTRSGGDLTLGLPSEEA PRSPLAPSEG 1080
Db 1021 EYLVPQGFCDPAPGAGGVHRRSSSTRSGGDLTLGLPSEEA PRSPLAPSEG 1080
Qy 1081 AGSDVFGDLGMAAKGLQSLPTDHPSPLOYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFGDLGMAAKGLQSLPTDHPSPLOYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPQPPSPRGPLPAARPAAGATLERAKTLPSPKNGVVKDVPFAGGAVENPEYLTPO 1200

Db 1141 NQDVRPQPPRSGPLPAAPAGATLPRKTLSPGKNGVVDVPAFGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAFPAFNLNYYWQDPPRGGAPSTFKGTPTAENPEYLGDDVPV 1255
Db 1201 GGAAPQHPHPPAFPAFNLNYYWQDPPRGGAPSTFKGTPTAENPEYLGDDVPV 1255
RESULT 12
AAE20479
ID AAE20479 standard; Protein; 1255 AA.
XX AAE20479;
AC AAE20479;
XX 01-JUL-2002 (first entry)
DT Human Her-2/neu protein.
DE Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
KW human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX Homo sapiens.
XX Key Location/Qualifiers
FT Region 1021..1030
FT /note= "Naturally processed HLA-B44-restricted epitope"
XX WO200214503-A2.
XX 21-FEB-2002.
XX 14-AUG-2001; 2001WO-US41733.
XX 14-AUG-2000; 2000US-225152P.
XX 28-SEP-2000; 2000US-236428P.
XX 21-FEB-2001; 2001US-270520P.
XX (CORI-) CORIXA CORP.
XX Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
PI McNeill PD, Vedwick TS;
XX WPI; 2002-280758/32.
DR N-PSDB; AAD32743.
XX Novel isolated Her-2/Neu polypeptide composition useful for therapy,
PT prevention and diagnosis of cancer, preferably breast cancer
XX Disclosure; Page 114-117; 129pp; English.
XX The invention relates to an isolated Her-2/Neu polypeptide composition
CC effective for eliciting an immune response. The invention is useful for
CC eliciting an immune response in a patient, where the patient is human
CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
CC The composition is useful for the therapy and diagnosis of cancer,
CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
CC and other compositions for the diagnosis, prevention and treatment of
CC human malignancies, for stimulating and/or expanding T cells specific for
CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a
CC patient. The invention is useful for stimulating a T cell response in a
CC human patient, as probe or primer for nucleic acid hybridisation, to
CC selectively form duplex molecules with complementary stretches of the
CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full
CC length gene from a suitable library, and to direct expression of a
CC polypeptide in appropriate host cells. The composition is useful in
CC prophylactic or therapeutic applications and for the treatment of cancer,
CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-
CC associated malignancies. The invention is useful in gene therapy. The
CC present sequence is human Her-2/neu protein.
XX Sequence 1255 AA;

Query Match 98.8%; Score 6729; DB 23; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLALLPPGAASQVCTGTDKMLRLPASPTHLDMLRHLVGGCQVQGNL 60
Db 1 MELAALCRWGLLALLPPGAASQVCTGTDKMLRLPASPTHLDMLRHLVGGCQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEQVQYIKANSKFIGITELQRLRIVRGTOLEFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEQVQYILIAHNQVRQVPLQRLRIVRGTOLEFEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTTLWKDIFHNQOLA 180
Db 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTTLWKDIFHNQOLA 180
Qy 181 LTLIDTNRSRACHPCSPMCKGSRGWSSSDCOSLTRVCAGGCARCKGPLPTDCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMCKGSRGWSSSDCOSLTRVCAGGCARCKGPLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPPLHNOEVTAEADGTQCEKSKPCARVCYGLGMEHLREVRAVTSAN 360
Db 301 YNYLSTDVGSCTLVCPPLHNOEVTAEADGTQCEKSKPCARVCYGLGMEHLREVRAVTSAN 360
Qy 361 IQEPAGCKIFGSLAFLPESFDGDPASNTAPLOEQLOVPETLEETGYLYISAWPDSLP 420
Db 361 IQEPAGCKIFGSLAFLPESFDGDPASNTAPLOEQLOVPETLEETGYLYISAWPDSLP 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLGISWGLRSLRSLRGLALIHNTHLFCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLGISWGLRSLRSLRGLALIHNTHLFCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRDECEVGEGLACHOLCARGHCWGPGPTQVCNCSQFLRGOEC 540
Db 481 PWDQLFRNPHQALLHTANRDECEVGEGLACHOLCARGHCWGPGPTQVCNCSQFLRGOEC 540
Qy 541 VECEVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFCVARC 600
Db 541 VECEVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFCVARC 600
Qy 601 PSGVKPDLNYPKWKPFDESGACQPCPINCTHSCVDLDDKGCAPAEQASPLTISVAVG 660
Db 601 PSGVKPDLNYPKWKPFDESGACQPCPINCTHSCVDLDDKGCAPAEQASPLTISVAVG 660
Qy 661 ILLVVLGVVFGILIKRQOKIRKYTNRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVVLGVVFGILIKRQOKIRKYTNRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780
Qy 781 YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNKRGLSGDILLNWCQIAGMSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNKRGLSGDILLNWCQIAGMSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILRRFT 900
Qy 901 HQSDVMSYGVTVWELMTFGAPYDGIIPAREIPDLLEKGERLPPOPICTIDYVMIMVKCW 960
Db 901 HQSDVMSYGVTVWELMTFGAPYDGIIPAREIPDLLEKGERLPPOPICTIDYVMIMVKCW 960
Qy 961 IDSECRPRFRELVSFERSMARDQRFVIONEDLGPASPLDSTFYRSLLEDGMDGLVDA 1020
Db 961 IDSECRPRFRELVSFERSMARDQRFVIONEDLGPASPLDSTFYRSLLEDGMDGLVDA 1020
Qy 1021 EYLVFPQQGFCDPPAPGAGMVHRRHSSTRSGGDLTLGLEPSEEAAPRPLAPSEG 1080

Db 1021 BEYLVPOQFFCPDPAPGAGGVHRRSSSTRSGGDLTLGLRPSBEEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGLMGAAKGLQSLPTDHPSPLOQRYSEDPTVPLRSETDGYVAPITCSPOPEYV 1140
Db 1081 AGSDVFDGLMGAAKGLQSLPTDHPSPLOQRYSEDPTVPLRSETDGYVAPITCSPOPEYV 1140
Qy 1141 NOPDVRPOPPSRGGLPAARPAAGATLERAKTSLPGKNGVVKDVFAGGAVENPEYLTTPQ 1200
Db 1141 NOPDVRPOPPSRGGLPAARPAAGATLERAKTSLPGKNGVVKDVFAGGAVENPEYLTTPQ 1200
Qy 1201 GGAAPQHPHPPAFSPADNLNLYWQDPPRGAPPTFKGTPTAENPEYLGDDVPV 1255
Db 1201 GGAAPQHPHPPAFSPADNLNLYWQDPPRGAPPTFKGTPTAENPEYLGDDVPV 1255

RESULT 13

AAM51143
ID AAM51143 standard; Protein; 1255 AA.

AC AAM51143;

DT 17-JUN-2002 (first entry)

XX Human Her-2/neu oncogene-encoded p185 glycoprotein.

XX Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
KW tyrosine kinase; receptor; c-erbB2; gene therapy.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Domain 1..653 "extracellular domain"

FT Domain 676..1255 "note= \"intracellular domain\""

FT Domain 990..1255 "note= \"phosphorylation domain\""

FT WO200212341-A2.

XX 14-FEB-2002.

XX 03-AUG-2001; 2001WO-US24283.

XX 03-AUG-2000; 2000US-0632507.

XX (CORI-) CORIXA CORP.

XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Cheever MA, Gheysen D;

XX WPI; 2002-241743/29.

XX N-PSDB; ABA92250.

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting
PT or enhancing an immune response to the protein, has Her-2/neu
PT extracellular domain fused to Her-2/neu intracellular or
PT phosphorylation domain
XX
PS Claim 68; Fig 7; 141pp; English.

XX The present sequence is that of human Her-2/neu (p185 glycoprotein
CC or c-erbB2), an oncogenic self-protein and target for anti-cancer
CC vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
CC in a variety of cancers, including breast, ovarian, colon, lung and
CC prostate cancer. Her-2/neu is a member of the tyrosine kinase
CC family of receptor-like glycoproteins. It comprises an extracellular
CC domain with homology to the epidermal growth factor receptor
CC (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
CC intracellular domain that also shows homology to EGFR. Its
CC overexpression correlates with a poor prognosis in breast and
CC ovarian cancers. The invention provides Her-2/neu fusion
CC proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In
CC preferred fusion proteins, the extracellular domain of a Her-2/neu
CC protein is fused to a Her-2/neu intracellular domain or
CC phosphorylation domain (or its deltaPDP fragment). An immune
CC response to Her-2/neu protein is elicited or enhanced by
CC administering the fusion protein in the form of a vaccine, or by
CC transfecting cells of an animal ex vivo with a nucleic acid
CC encoding the fusion protein, and delivering the transfected cells
CC to the animal. The fusion proteins, nucleic acids, and isolated
CC specific T-cells are useful for inhibiting the development of a
CC cancer, especially breast, ovarian, colon, lung or prostate cancer
CC in a patient. T cells that specifically react with a Her-2/neu
CC fusion protein can be used to remove tumour cells from a sample in
CC order to inhibit the development of cancer in a patient.

XX Sequence 1255 AA;

Query Match 98.8%; Score 6729; DB 23; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLALLPPGAASSTQVCTGDMKRLPASPETHLDMRLHLYGCGVQGNL 60
Db 1 MELAALCRWGLLALLPPGAASSTQVCTGDMKRLPASPETHLDMRLHLYGCGVQGNL 60
Qy 61 ELTYLPTNASISFLQDIQEVQYIKANSKFTIGITELQRLRIVRGTLQFEDNYALAVLDNG 120
Db 61 ELTYLPTNASISFLQDIQEVQYIKANSKFTIGITELQRLRIVRGTLQFEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVGTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHNQOLA 180
Db 121 DPLNNTTPTVGTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHNQOLA 180
Qy 181 LTLIDNTRSRACHPCSPMKGSCWGESSEDCQSLRTVTCAGGCARCKGPLPTCCHEQC 240
Db 181 LTLIDNTRSRACHPCSPMKGSCWGESSEDCQSLRTVTCAGGCARCKGPLPTCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPLNHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREVAVTSAN 360
Db 301 YNYLSTDVGSCTLVCPLNHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREVAVTSAN 360
Qy 361 IQEPAGCKKIFGSLAFIPESFDGDPASNTAPLQEQLOVFTLEETITGYLYISAWPDSLP 420
Db 361 IQEPAGCKKIFGSLAFIPESFDGDPASNTAPLQEQLOVFTLEETITGYLYISAWPDSLP 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLSLTQGLGISWLGRLSRLRELGLALIHNNHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLSLTQGLGISWLGRLSRLRELGLALIHNNHLCFVHTV 480
Qy 481 PWDOLFNPQHALLHTANRPEDECEVCGSLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
Db 481 PWDOLFNPQHALLHTANRPEDECEVCGSLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
Qy 541 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFPGPADQCVACAHYKDPFCVARC 600
Db 541 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFPGPADQCVACAHYKDPFCVARC 600
Qy 601 PSGVKPDLNMPYIMKFPDESGACQPCINCTHSCVDLDDKGCPEQASPLTSTVSAVVG 660
Db 601 PSGVKPDLNMPYIMKFPDESGACQPCINCTHSCVDLDDKGCPEQASPLTSTVSAVVG 660
Qy 661 ILLVVLGVVFGILI KRRQKIRKYTMRRLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVLGVVFGILI KRRQKIRKYTMRRLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Qy 721 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVAGVGSPP 780
Db 721 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVAGVGSPP 780

QY 781 YVSRLLGICLTSTVOLVTLQMPYGLDHRVNRGRGLSQDGLLNWCMQIAKGMVLEDDR 840
Db 781 YVSRLLGICLTSTVOLVTLQMPYGLDHRVNRGRGLSQDGLLNWCMQIAKGMVLEDDR 840
QY 841 LVHRDLAARNVLKSPNWKITDFGLARLLDIDETEHADGKVPKWKMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNWKITDFGLARLLDIDETEHADGKVPKWKMALESILRRFT 900
QY 901 HQSDVWSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICTIDVYIMVKCWM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICTIDVYIMVKCWM 960
QY 961 IDSECRPRFRELVSFSEMRADPQRFVVIQNEGLGPASPLDSTFYRSLLEDGDLVDA 1020
Db 961 IDSECRPRFRELVSFSEMRADPQRFVVIQNEGLGPASPLDSTFYRSLLEDGDLVDA 1020
QY 1021 EYLVPQOGFCPPDAPGAGGVHHRSSSTRSGGDLTLGLEPSEBEAPRSLAPSEG 1080
Db 1021 EYLVPQOGFCPPDAPGAGGVHHRSSSTRSGGDLTLGLEPSEBEAPRSLAPSEG 1080
QY 1081 AGSDVDFDGLGWAAGKGLQSLTPHDPSPLOYSEDPTVPLFSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVDFDGLGWAAGKGLQSLTPHDPSPLOYSEDPTVPLFSETDGYVAPLTCSPQPEYV 1140
QY 1141 NQDVRPQPPSPREGPLPAARPAGATLBRATLSPGKNGVVDVFAFGAVENPEYLTPO 1200
Db 1141 NQDVRPQPPSPREGPLPAARPAGATLBRATLSPGKNGVVDVFAFGAVENPEYLTPO 1200
QY 1201 GGAAPOPHPPAFSPAFDNLVYDODPPERGAPPSTFKGTAEENPEYLGDDVPV 1255
Db 1201 GGAAPOPHPPAFSPAFDNLVYDODPPERGAPPSTFKGTAEENPEYLGDDVPV 1255

RESULT 14

AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX AC AAU77114;
XX DT 05-JUN-2002 (first entry)
XX DE Human Her-2/neu polypeptide.
XX KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
XX KW Hodgkin's lymphoma; T cell therapy.
XX OS Homo sapiens.
XX FN WO200213847-A2.
XX PD 21-FEB-2002.
XX PF 13-AUG-2001; 2001WO-US25408.
XX PR 14-AUG-2000; 2000US-0638280.
XX PR 28-SEP-2000; 2000US-0675904.
XX PA (CORI-) CORIXA CORP.
XX GA Gaiger A, Cheever MA, Hand-zimmermann S;
XX WPI; 2002-280741/32.
XX DR N-PSDB; ABK10730.
XX PT Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
XX polypeptide.
PS Disclosure; Page 71-74; 74pp; English.
XX

CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX

SQ Sequence 1255 AA;

Query Match 98.8%; Score 6729; DB 23; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALLPPGAASQVCTGDKMLRASPETHLDMLRHLVGGCQVQGNL 60
Db 1 MELAALCRWGLLALLPPGAASQVCTGDKMLRASPETHLDMLRHLVGGCQVQGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQYIKANSKFTIGITELQRLRIVRGTLQFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQYVLIAHNQVRVPLQRLRIVRGTLQFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGTASPGGURELQLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNQOLA 180
Db 121 DPLNNTTPTVGTASPGGURELQLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNQOLA 180
QY 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLRTVCAGGCARCKGPLPTCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLRTVCAGGCARCKGPLPTCCHEQC 240
QY 241 AAGCTGPKHSDCLACLHFNHSGICEHLCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLHFNHSGICEHLCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQRCCKSPCARVCYGLGMHLEHREAVTSAN 360
Db 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQRCCKSPCARVCYGLGMHLEHREAVTSAN 360
QY 361 IQEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLOVFETLEETGYLIYSWPPSLP 420
Db 361 IQEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLOVFETLEETGYLIYSWPPSLP 420
QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLRSLRGLSLALIHNNHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLRSLRGLSLALIHNNHLCFVHTV 480
QY 481 PWDQLFNPHQALLHTANRPEDECVGEGLAHQLCARHGWGPGPTQCVNCSQFLRGQEC 540
Db 481 PWDQLFNPHQALLHTANRPEDECVGEGLAHQLCARHGWGPGPTQCVNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQCVCAHYKDPFCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQCVCAHYKDPFCVARC 600
QY 601 PSGVKPDLSPNPIWKFPDEEGACQPCINCHSCVDLDDKGCPEAQASPLTSTVSAVVG 660
Db 601 PSGVKPDLSPNPIWKFPDEEGACQPCINCHSCVDLDDKGCPEAQASPLTSTVSAVVG 660
QY 661 ILVVLGVVFGILIKRRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILVVLGVVFGILIKRRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAMRILKETEL 720
QY 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKETLDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKETLDEAYVMAGVGP 780
QY 781 YVSRLLGICLTSTVOLVTLQMPYGLDHRVNRGRGLSQDGLLNWCMQIAKGMVLEDDR 840
Db 781 YVSRLLGICLTSTVOLVTLQMPYGLDHRVNRGRGLSQDGLLNWCMQIAKGMVLEDDR 840
QY 841 LVHRDLAARNVLKSPNWKITDFGLARLLDIDETEHADGKVPKWKMALESILRRFT 900

|||||
841 LVHRDLAARNVVKSPNHNKIDTGLARLLDIDETEHADGGKVPKWMALLESILRRFT 900
Qy
901 HQSDVWSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICITIDVYMIWVKCWM 960
Db
901 HQSDVWSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICITIDVYMIWVKCWM 960
Qy
961 IDSECRPRFRELVSFSEFMRDQRFVVIQNEIDLGPASPLDSTFYRSILLEDDMGDLVDA 1020
Db
961 IDSECRPRFRELVSFSEFMRDQRFVVIQNEIDLGPASPLDSTFYRSILLEDDMGDLVDA 1020
Qy
1021 BEYLVPQOGFFCPCPAPGAGGMVHRHRSSTRSGGDLTLGLSPSEEAAPRSLAPSEG 1080
Db
1021 BEYLVPQOGFFCPCPAPGAGGMVHRHRSSTRSGGDLTLGLSPSEEAAPRSLAPSEG 1080
Qy
1081 AGSDVFDGDLGNGAAKGLQSLPTHDPSPQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db
1081 AGSDVFDGDLGNGAAKGLQSLPTHDPSPQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy
1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVPFAGGAVENPEYLTPO 1200
Db
1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVPFAGGAVENPEYLTPO 1200
Qy
1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYILGDDVPV 1255
Db
1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYILGDDVPV 1255

RESULT 15

AAR39568

ID AAR39568 standard; Protein; 1433 AA.

XX

AC AAR39568;

XX

DT 07-FEB-1994 (first entry)

XX

DE Sequence of c-erbB-2 tumour antigen.

XX

KW Tumour antigen; c-erbB-2; glycoprotein.

XX

OS Homo sapiens.

XX

PN WO9316185-A.

XX

PD 19-AUG-1993.

XX

PF 05-FEB-1993; 93WO-US01055.

XX

PR 06-FEB-1992; 92US-0831967.

XX

PA (CETU) CETUS ONCOLOGY CORP.

XX

PA (CREA-) CREATIVE BIOMOLECULES INC.

XX

PI Houston LL, Huston JS, Oppermann H, Ring DB;

XX

DR WPI; 1993-272889/34.

XX

DR N-PSDB; AAQ46083.

XX

XX New single chain Fv polypeptide binding to C-erbB-2 tumour

PT antigen - for imaging or treating breast or ovarian cancer etc.

XX

PS Disclosure; pages 48-54; 87pp; English.

XX

CC c-erbB-2 refers to a protein antigen expressed on the surface of

CC tumour cells. such as breast and ovarian tumour cells, which is an

CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric

CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents

CC the location of a stop codon in AAQ46083.

XX

SQ Sequence 1433 AA;

XX

Query Match 98.2%; Score 6686; DB 14; Length 1433;

Best Local Similarity 98.2%; Pred. No. 0;

Matches 1233; Conservative 5; Mismatches 17; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLALLPPGAASQVCTGTDMLRLPASPTHDLMLRHLHLYOGCOVVOGNL 60
Db
1 MELAALCRWGLLALLPPGAASQVCTGTDMLRLPASPTHDLMLRHLHLYOGCOVVOGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQYIKANSKPIGTELQRLRIVRGTQOLFEDNYALAVLDNG 120
Db
61 ELTYLPTNASLSFLQDIQEVQYVLIHQNQVQPLQRLRIVRGTQOLFEDNYALAVLDNG 120
Qy 121 DPLNTPPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTTLWKDIFHKNOLA 180
Db
121 DPLNTPPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTTLWKDIFHKNOLA 180
Qy 181 LTLIDTNRSPACHPCSPMKSGSCWSESSDCSLTRTVCCAGCARCKGPLPTDCCHEQC 240
Db
181 LTLIDTNRSPACHPCSPMKSGSCWSESSDCSLTRTVCCAGCARCKGPLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db
241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPLNHQNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVAVTSAN 360
Db
301 YNYLSTDVGSCTLVCPLNHQNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVAVTSAN 360
Qy 361 IQEFAGCKKIFGSLAFIPESFDGPASNTAPLQEQLOVFETLEEITGYLYISAWPDSLP 420
Db
361 IQEFAGCKKIFGSLAFIPESFDGPASNTAPLQEQLOVFETLEEITGYLYISAWPDSLP 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSLRELGLSLALHNNTHLSFVHTV 480
Db
421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSLRELGLSLALHNNTHLSFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQICARGHCWGPGPTQCVNCSQFLRGQEC 540
Db
481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQICARGHCWGPGPTQCVNCSQFLRGQEC 540
Qy 541 VEECRVLOGLPREYVNAHCLPCHPEQOPONGSVTCFGEADOCVACAHYKDPFCVARC 600
Db
541 VEECRVLOGLPREYVNAHCLPCHPEQOPONGSVTCFGEADOCVACAHYKDPFCVARC 600
Qy 601 PSGVKPDLSPYMPWKFPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSISAVVG 660
Db
601 PSGVKPDLSPYMPWKFPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSISAVVG 660
Qy 661 ILLVVLGVVFGILIKRRQKIRKYTMRLLEQTELVEPLTPSGAMPNQAMRILKETEL 720
Db
661 ILLVVLGVVFGILIKRRQKIRKYTMRLLEQTELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRNTSPKANKETILDEAYVWAGVSP 780
Db
721 RKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRNTSPKANKETILDEAYVWAGVSP 780
Qy 781 YVSRLLIGICLTSTVQLVTLMPYGLLDHVRNRLGSLQDLNWCMIQAKGMSYLEVDR 840
Db
781 YVSRLLIGICLTSTVQLVTLMPYGLLDHVRNRLGSLQDLNWCMIQAKGMSYLEVDR 840
Qy 841 LVHRDLAARNVVKSPNHNKIDTGLARLLDIDETEHADGGKVPKWMALLESILRRFT 900
Db
841 LVHRDLAARNVVKSPNHNKIDTGLARLLDIDETEHADGGKVPKWMALLESILRRFT 900
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICITIDVYMIWVKCWM 960
Db
901 HQSDVWSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICITIDVYMIWVKCWM 960
Qy 961 IDSECRPRFRELVSFSEFMRDQRFVVIQNEIDLGPASPLDSTFYRSILLEDDMGDLVDA 1020
Db
961 IDSECRPRFRELVSFSEFMRDQRFVVIQNEIDLGPASPLDSTFYRSILLEDDMGDLVDA 1020
Qy 1021 BEYLVPQOGFFCPCPAPGAGGMVHRHRSSTRSGGDLTLGLSPSEEAAPRSLAPSEG 1080
Db
1021 BEYLVPQOGFFCPCPAPGAGGMVHRHRSSTRSGGDLTLGLSPSEEAAPRSLAPSEG 1080

```
Qy 1081 AGSDVFDGDLGWAAGKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSQPPEYV 1140
    |||||
Db 1081 AGSDVFDGDLGWAAGKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSQPPEYV 1140
    |||||
Qy 1141 NOPDVRPQPPSPRSGPLPAARPAGATLERAKTSLSGKNGVVKDVFAGGAVENPEYLTPO 1200
    |||||
Db 1141 NOPDVRPQPPSPRSGPLPAARPAGATLERAKTSLSGKNGVVKDVFAGGAVENPEYLTPO 1200
    |||||
Qy 1201 GGAAPQHPHPPAFSPFDNLYYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
    |||||
Db 1201 GGAAPQHPHPPAFSPFDNLYYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
    |||||
```

Search completed: July 22, 2003, 08:40:17
Job time : 42.9774 secs

THIS PAGE BLANK (USPTO)